

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ  
Государственное образовательное учреждение высшего профессионального образования  
«ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

---

**В.П. Комагоров**

# **ТЕХНОЛОГИИ СЕТИ ИНТЕРНЕТ: ПРОТОКОЛЫ И СЕРВИСЫ**

*Рекомендовано в качестве учебного пособия  
Редакционно-издательским советом  
Томского политехнического университета*

Издательство  
Томского политехнического университета  
2009

УДК 681.324

К63

**Комагоров В.П.**

К63 Технологии сети Интернет: протоколы и сервисы: учебное пособие / В.П. Комагоров; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2009. – 107 с.

Пособие содержит сведения о сервисах сети Интернет и сетевых протоколах, которые обеспечивают поддержку этих сервисов. В конце каждого раздела пособия приведены методические указания, отражающие ключевые моменты этого раздела.

Пособие подготовлено на кафедре оптимизации систем управления ТПУ в рамках инновационной образовательной программы «Информационно-телекоммуникационные системы и технологии» в части обеспечения магистерской программы «Сети ЭВМ и телекоммуникации».

**УДК 681.324**

*Рецензенты*

Кандидат технических наук, доцент  
зам. заведующего кафедрой оптимизации  
систем управления ТПУ  
*О.Б. Фофанов*

Кандидат технических наук, доцент  
научный руководитель  
НО «Некоммерческий фонд развития региональной энергетики»  
*В.Г. Ротарь*

Кандидат технических наук,  
доцент кафедры компьютерных систем  
в управлении и проектировании ТУСУРа  
*В.П. Коцубинский*

© Томский политехнический университет, 2009

© Комагоров В.П., 2009

© Оформление. Издательство Томского  
политехнического университета, 2009

## ВВЕДЕНИЕ

Глобальная сеть Internet содержит информационные ресурсы по различным направлениям человеческой деятельности: науке, производству, культуре, спорту. Оказать помощь в получении знаний для эффективной работы с этими ресурсами является целью настоящего пособия.

Пособие разрабатывалось в рамках инновационной образовательной программы «Информационно-телекоммуникационные системы и технологии» в части обеспечения магистерской программы «Сети ЭВМ и телекоммуникации» и содержит десять глав.

В первой главе студенты знакомятся с понятием сервиса Internet, видами сервисов, сетевыми протоколами, обеспечивающими их поддержку. Эти сведения базируются на материале, изложенном в [1, 2].

Вторая глава содержит описание системы доменных имен (системы DNS). При изложении этой главы были использованы материалы, содержащиеся в [1, 2].

Третья глава посвящена сервису Telnet. Основным источником для описания этого сервиса явились материалы, изложенные в [1, 3].

В четвертой главе рассмотрен сервис FTP. Глава содержит описание протокола FTP, который является основой этого сервиса. Материалы главы заимствованы из [1].

Пятая глава посвящена сервису Usenet – системе телеконференций (группам новостей). Более подробно эти материалы изложены в [4, 5].

В шестой главе описан сервис E-mail – электронная почта. Основой для этой главы явились материалы, представленные в [1].

Седьмая глава содержит описание сервиса WWW – гипертекстовой системы интеграции сетевых ресурсов в единое информационное пространство. Эти материалы заимствованы из [1].

Восьмая глава посвящена сервису SE – поисковым системам. В ней приводится описание принципов работы поисковых систем Яндекс, Google и Rambler. Источниками этих материалов являются [6, 7, 8, 9].

В девятой главе рассмотрен сервис IRC (*чат*) – системы для общения через Internet в режиме реального времени. Для этого раздела были использованы материалы, изложенные в [10, 11].

Десятая глава содержит описание сервиса RTVC – видеоконференций реального времени. Материалы этой главы заимствованы из [12, 13, 14].

В конце каждой главы приведены методические указания, содержащие основные выводы по главе.

# ГЛАВА 1. ПРОТОКОЛЫ И СЕРВИСЫ INTERNET

## 1.1. Понятие сервиса Internet

В глобальной сети Internet функционирует ряд программ и сетевых протоколов прикладного уровня, которые позволяют пользователю сети получать удаленный доступ к информационным ресурсам, вести активный диалог с другими удаленными пользователями, отправлять и получать почту.

Для организации работы пользователей сети с распределенными информационными ресурсами, необходимы следующие компоненты:

- программа, установленная на компьютере пользователя, которая может осуществлять сетевой запрос с целью получения информационного объекта, и предназначенная для его обработки;
- программа, установленная на компьютере, где расположен информационный объект, которая может осуществлять по сетевому запросу поиск и пересылку объекта, а также упорядочивание доступа к нему нескольких пользователей;
- сетевой протокол взаимодействия между этими программами.

Технология взаимодействия, в которой одна программа запрашивает выполнение какой-либо совокупности действий (услугу), а другая ее выполняет, называется *технологией «клиент-сервер»*. Участники такого взаимодействия называются соответственно *клиентом (client)* и *сервером (server)*. Достаточно часто клиентом или сервером называют компьютеры, на которых функционирует клиентское или серверное программное обеспечение.

*Комплекс программ и протоколов прикладного уровня, обеспечивающий пользователю сети Internet возможность работы с распределенными информационными ресурсами, принято называть сервисами.*

К числу основных сервисов сети Internet относятся:

- *DNS ((Domain Name System)*. Система доменных имен, позволяющая использовать для адресации узлов сети мнемонические имена вместо числовых адресов.
- *Telnet*. Обеспечивает удаленный доступ к серверам сети Internet. С помощью программы-клиента Telnet можно войти в систему удаленного компьютера в качестве его терминала и осуществить исполнение команд программного обеспечения сервера (например, соединиться с другим сервером, получить доступ к данным).
- *FTP (File Transfer Protocol)*. Передача файлов между соединенными посредством сети компьютерами. Этот сервис позволяет уста-

навливать логическую связь с удаленным узлом, осуществлять поиск необходимого файла в каталогах, выполнять прием и передачу файлов, создавать собственные файлы и каталоги и удалять их.

- *Usenet*. Телеконференции по Internet. Это система организации публичных обсуждений по конкретным темам: вычислительная техника, социальные вопросы, новости и т. д. Для организации телеконференций существует список серверов телеконференций.
- *E-mail*. Электронная почта. Это система пересылки сообщений между людьми, имеющими доступ к сети Internet. Для передачи почтовых сообщений используется протокол SMTP. Доставка почты пользователю из почтового ящика выполняется протоколом Post Office Protocol (POP). В системе реализованы такие функции как: создать сообщение, доставить почту, переслать полученное письмо, уведомить о получении, уведомить о прочтении.
- *World Wide Web (WWW)*. Доступ к гипертекстовым страницам WWW-серверов. Основными носителями информации в сети Internet являются мультимедийные WWW-сервера, на которых информация размещается в виде Web-страниц. Эти страницы отображают звук, изображение (рисунок, фотографию т. д.) и гипертекст. Для доступа к WWW-серверам используются специальные программы (browser): Internet Explorer, Netscape Navigator и др.

Перечисленные сервисы относятся к стандартным. Это означает, что принципы построения клиентского и серверного программного обеспечения, а также протоколы их взаимодействия соответствуют международным стандартам.

Наряду со стандартными сервисами существуют и нестандартные, представляющие собой оригинальные разработки различных компаний. Наиболее значимыми из них являются:

- *SE (Search Engine)*. Поисковая система (машина). Это система поиска необходимой информации в сети Internet.
- *IRC (Internet Relay Chat)*. Телеконференции реального времени (чат – соединение). Чатом (chat) называется общение между пользователями с помощью клавиатуры компьютера. При этом в чате могут участвовать одновременно несколько пользователей, подключенных к серверам, работающим в режиме многоканального чата.
- *RTVC (Real Time Video Connection)*. Видеоконференции реального времени. Видеоконференция – это взаимное общение двух или более лиц с использованием мультимедийных возможностей компьютеров по сети Internet. Компьютер, подключенный к сети, должен быть оснащен видеокамерой, микрофоном и аудиоконками.

## 1.2. Протоколы сервисов Internet

Для стандартных сервисов также стандартизируется и интерфейс взаимодействия с протоколами транспортного уровня TCP и UDP, описание которых содержится в [2]. В частности, за каждым сервером резервируются стандартные номера TCP- и UDP-портов, которые остаются неизменными независимо от особенностей той или иной фирменной реализации как компонентов сервиса, так и транспортных протоколов (табл. 1).

Таблица 1

*Сервисы и стандартные номера TCP- и UDP-портов*

Компонент сервиса	Номер порта	Протоколы транспортного уровня
DNS		
DNS-сервер	53	TCP, UDP
Telnet		
Telnet-сервер	23	TCP
FTP		
FTP-сервер	20, 21	TCP
Usenet		
NNTP-сервер	119	TCP
E-mail		
SMTP-сервер	25	TCP
POP3-сервер	110	TCP
IMAP-сервер	143	TCP
WWW		
HTTP-сервер	80	TCP

Номера портов клиентского программного обеспечения так жестко не регламентируются. Это объясняется следующими факторами:

- во-первых, на пользовательском узле может функционировать несколько копий клиентской программы, и каждая из них должна однозначно идентифицироваться транспортным протоколом, т. е. за каждой копией должен быть закреплен свой уникальный номер порта;
- во-вторых, клиенту важна регламентация портов сервера, чтобы знать, куда направлять запрос, а сервер сможет ответить клиенту, узнав адрес из поступившего запроса.

Для нестандартных сервисов, представляющих собой оригинальные разработки различных компаний, номера TCP- и UDP-портов сервера и клиентской программы не регламентируются.

## Методические указания

При изучении этого раздела необходимо усвоить следующие понятия:

- технология взаимодействия, в которой одна программа запрашивает выполнение какой-либо совокупности действий (услугу), а другая ее выполняет, называется технологией «клиент-сервер»;
- комплекс программ и протоколов прикладного уровня, обеспечивающий пользователю сети Internet возможность работы с распределенными информационными ресурсами, принято называть сервисами;
- к числу основных (стандартных) сервисов сети Internet относятся: DNS – система доменных имен, Telnet – управление удаленными компьютерами в терминальном режиме, FTP – система файловых архивов, Usenet – система телеконференций, E-mail – электронная почта, WWW – гипертекстовая система интеграции сетевых ресурсов в единое информационное пространство;
- наряду со стандартными сервисами существуют и нестандартные, представляющие собой оригинальные разработки различных компаний: SE – поисковые системы, IRC – телеконференции реального времени (чат – соединение), RTVC – видеоконференции реального времени.
- для стандартных сервисов за каждым сервером резервируются стандартные номера TCP- и UDP-портов, номера портов клиентского программного обеспечения жестко не регламентируются;
- для нестандартных сервисов номера TCP- и UDP-портов сервера и клиентской программы не регламентируются.

## ГЛАВА 2. СЕРВИС DNS – СИСТЕМА ДОМЕННЫХ ИМЕН

### 2.1. Назначение сервиса DNS

IP-адрес в сети Internet составляет 32 бита в том случае, когда используется протокол IPv4, и 128 бит для протокола IPv6. Представление IP-адреса в двоичном виде не удобно для запоминания пользователем. Поэтому для задания IP-адресов широко используется *система имен доменов (Domain Name System, DNS)*.

*Термин «домен» определяется как сфера деятельности, отношений или выполнения каких-либо совместных функций.*

Система DNS позволяет обращаться к сетевым компьютерам не только по их IP-адресам, но и по индивидуальным адресам, например, вместо IP-адреса в двоичном виде 11000000 01100110 11111001 00000011 или «десятичное с точкой» 192.102.249.3 можно задать имя компьютера SONET.com.

Каждое из составляющих имени, например, *ftp.microsoft.com*, называется *меткой*. В данном случае имя компьютера состоит из трех меток: *ftp*, *microsoft* и *com*. Метки отделены друг от друга точками.

Метка *ftp* означает, что данный компьютер является *хостом ftp*, то есть на нем работает *ftp-сервер*. Метка *microsoft* описывает организацию (сферу деятельности), которой принадлежит компьютер корпорации Microsoft. Метка *com* обозначает, что данный компьютер выполняет коммерческие функции.

Таким образом, DNS выполняет две основных функции:

- организация иерархического пространства имен;
- обеспечение разрешения (т. е. поиска соответствия) доменных имен и IP-адресов.

Как и большинство сервисов, DNS относится к средствам прикладного уровня и строится по принципу «клиент-сервер».

В структуре сервиса DNS выделяют следующие компоненты:

- *Информационный ресурс* – иерархически организованное пространство доменных имен. Соответствия доменных имен и IP-адресов описывается в распределенной по специальным узлам сети, называемым серверами имен, базе данных. Часть иерархического пространства имен, обслуживаемая одним сервером имен и представленная в его локальной базе данных, называется *зоной ответственности (zone of authority)*.
- *DNS-клиент (resolver)* – программный модуль, который обеспечивает выполнение запросов к серверу имен с целью разрешения до-

менного имени. Как правило, DNS-клиент входит в состав операционной системы.

- *Сервер имен (name server)*, или *DNS-сервер*, – программа, обеспечивающая хранение части распределенной базы данных соответствий IP-адресов и доменных имен, а также осуществляющая по запросу клиента поиск IP-адреса на основе предложенного доменного имени.
- *Протокол DNS* – протокол взаимодействия DNS-клиентов и DNS-серверов.

Следует отметить, что часто терминами DNS-клиент и DNS-сервер обозначают компьютеры, на которых функционирует соответственно клиентское или серверное программное обеспечение.

## 2.2. Организация пространства имен

Пространство доменных имен имеет иерархическую структуру. На самом верхнем уровне иерархии располагается корневой домен. Следующий уровень иерархии составляют домены верхнего или первого. Каждый домен верхнего уровня включает в себя домены второго уровня и т. д.

Основные принципы построения системы DNS изложены в [2]. В соответствии с этим отметим некоторые аспекты ее дальнейшего развития.

Теоретически домен любого уровня может содержать в себе как отдельные узлы, представленные своими именами, так и домены более низкого уровня (субдомены). Однако, на практике домены, уровень которых ниже третьего, встречаются крайне редко.

Домены первого уровня делятся на три группы:

- домены общего назначения;
- национальные домены;
- обратный домен.

Первую группу составляют *домены общего назначения* (табл. 2). Изначально домены общего назначения предназначались для объединения доменов нижних уровней, принадлежащих организациям и учреждениям США. Поэтому в силу традиции большая часть доменов, зарегистрированных за организациями других государств, входят не в домены общего назначения, а в так называемые национальные домены. Однако ничто не мешает какой-нибудь компании, например, Российской, зарегистрировать домен второго уровня в домене «com». Исключением являются только домены «mil» и «gov», которые используются только учреждениями и организациями США.

Другой традицией для доменов общего назначения до последнего времени было жесткое ограничение на длину имени – три символа. Ввод в 2001 году доменов «name», «info» и «museum» нарушил эту

традицию. По всей видимости, это не будет соблюдаться и в дальнейшем, поскольку в ближайшее время среди прочих планируется ввести еще два домена общего назначения с «неправильным» именем – «соор» и «аего».

Таблица 2

*Домены общего назначения*

Название домена	Комментарий
«Старые» домены, созданные в 1984 году	
«com»	COMmercial, коммерческие организации
«gov»	GOVERNment, правительственные учреждения США
«int»	INTernational Organizations, международные организации
«mil»	MILitary, военные организации США
«edu»	EDUcational, образовательные проекты и учреждения
«org»	ORGanisations, некоммерческие организации или организации, не попадающие в другие категории
«net»	NETwork, сети общего назначения
«Новые» домены, созданные в 2001 году	
«info»	INFOrmation, домен свободного использования для предоставления информации в Интернет
«biz»	Business Organizations, различные организации
«name»	Домен предназначен для использования частными лицами
«museum»	Музеи

Во вторую группу включены *национальные домены*. Имя каждого такого домена состоит из *двух символов* и представляет собой сокращение названия государства (так называемый «код страны»), которому принадлежит домен, например, «ru» означает Россия. Список национальных доменов разработан и утвержден Национальным Институтом Стандартов США (ISO 3166-1).

Третья группа состоит из одного домена с четырехсимвольным именем «агра», предназначенного для поиска доменного имени по IP-адресу (обратного разрешения). Часто этот домен называют обратным доменом (reverse domain).

Каждый домен верхнего уровня, как правило, включает в себя домены второго уровня, имена которых выбираются относительно произвольно, например, по имени организации, за которой зарегистрировано это имя, или по названию региона. Порядок создания доменов второго уровня определяется администраторами соответствующего родительского домена верхнего уровня.

Так, например, в национальном домене первого уровня «ru» установлен порядок, согласно которому внутри этого домена выделяются три группы доменов второго уровня.

1. Домены общего пользования («домены типа GENERIC»):
  - «ac.ru» – домен Академии Наук России;
  - «com.ru» – коммерческие организации;
  - «edu.ru» – образовательные проекты и организации;
  - «net.ru» – сети, принадлежащие различным организациям;
  - «org.ru» – некоммерческие организации;
  - «int.ru» – домен для общего использования;
  - «mil.ru» – военные организации и учреждения;
  - «pp.ru» – домен для использования частными лицами.
2. Домены субъектов Российской Федерации («домены типа GEOGRAPHICAL»).
3. Другие домены, имена которых выбираются произвольным образом, например, по имени владельца (имени человека, названию организации), обобщенной тематике, которой соответствует информация, представленная на узлах домена и т. п. Например, существует домен второго уровня «fio.ru», закрепленный за Федерацией Интернет Образования.

Аналогично доменам второго уровня структуру доменов более низких уровней определяет администрация родительского домена. Поэтому не существует какой-либо единой для всех схемы структуризации таких доменов.

Рассмотрим в качестве примера организацию домена Федерации Интернет Образования (рис. 1). Федерация имеет домен второго уровня «fio», зарегистрированный в домене «ru». В данном домене представлены как отдельные узлы, например, узел с именем «www» – узел, на котором размещен Web-сайт Федерации, информационные узлы различных проектов Федерации («parent», «teacher», «writer» и т. д.), так и различные субдомены третьего уровня, например, домен «center» – домен Московского Центра Интернет-образования, домен «net» – внутренняя сеть Федерации, домен «dlmsk» – домен поддержки системы дистанционного образования Московского Центра, а также домены региональных центров Федерации, названия которых определяются регионами (например, «spb» или «samara»). В свою очередь, домен третьего уровня «net» содержит в себе субдомен «msk» – домен, в котором размещены компьютеры учебных классов Московского Центра.

Общие правила построения имен доменов и узлов следующие: во-первых, имя может состоять только из букв латинского алфавита, цифр и символа «-» (дефис), а, во-вторых, длина имени не может превышать 63 символов.

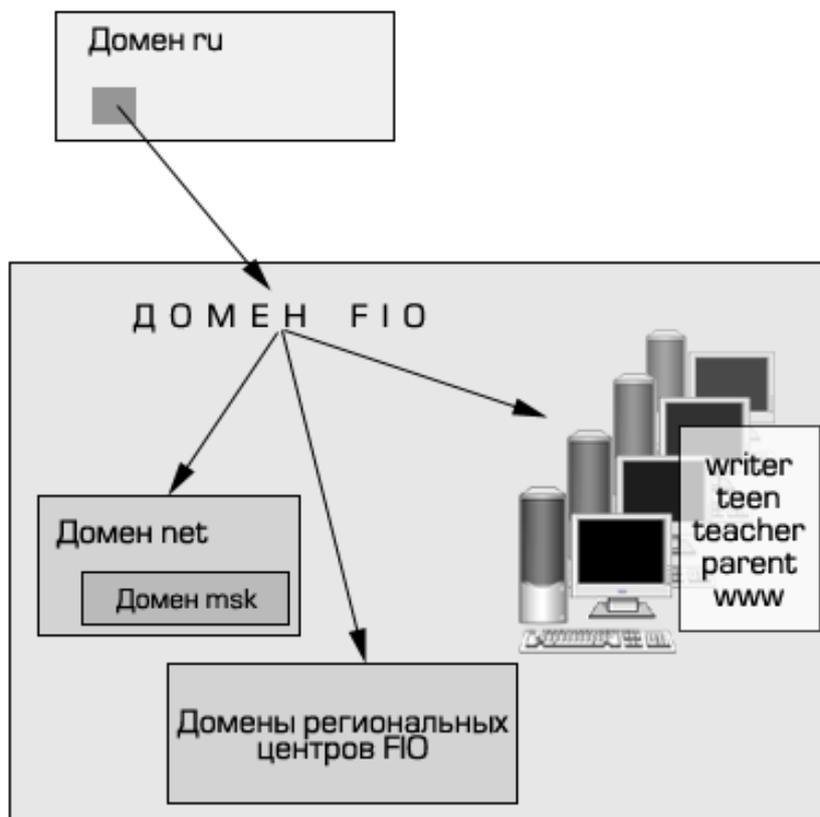


Рис. 1. Домен Федерации Интернет Образования

Кроме того, доменные имена являются нечувствительными к регистру символов, входящих в его состав. Это означает, что последовательности символов «Com», «COM», «cOm», «com» и т. п. обозначают одно и то же имя.

Для адресации конкретного узла или домена используется так называемое *полностью определенное доменное имя (Fully Qualified Domain Name, FQDN)*. FQDN – это имя домена или узла, отражающее всю иерархию включения данного домена или узла в домены более высокого уровня вплоть до корневого домена. Полное доменное имя строится справа налево и последовательно включает в себя имена всех доменов иерархии, разделенные точками, начиная с первого уровня и заканчивая доменом или узлом, которому это имя принадлежит. Например, для узла «www», принадлежащего домену «fio», полное доменное имя будет иметь вид «www.fio.ru».

Следует отметить, иерархическая организация пространства имен не накладывает никаких ограничений на соответствие доменных имен физическим узлам сети. Поэтому один и тот же узел сети может входить в различные домены, причем разного уровня, кроме того, узел может иметь несколько имен в одном домене. В таких случаях он будет иметь различные полностью определенные доменные имена.

Например, узел 213.128.193.116 входит под именем «www» в домен 3-его уровня «center.fio.ru», а также под таким же именем «www» в домен 4-го уровня «msk.net.fio.ru». Таким образом, он имеет два полностью определенных имени: «www.center.fio.ru» и «www.msk.net.fio.ru». В качестве другого примера можно привести узел 213.128.193.115, имеющий несколько имен в домене «msk.net.fio.ru», в том числе имена «msk-server.msk.net.fio.ru» и «mail.msk.net.fio.ru».

В настоящее время активно разрабатывается проект IDN (Internationalized Domain Name, «Интернационализованное» доменное имя), согласно которому в доменные имена могут входить символы любых национальных алфавитов. В настоящее время ведется тестовая регистрация имен с использованием символов китайского, корейского и японского алфавитов в доменах «com», «net» и «org», а также – символов русского алфавита в домене «ru». Однако, поскольку пока не принят официальный стандарт IDN как расширение классической системы DNS, корректное разрешение доменных имен, включающих в себя национальные символы, не гарантируется.

### 2.3. База данных сервера DNS

База данных сервера DNS предназначена для описания соответствий доменных имен IP-адресам узлов.

Важным преимуществом DNS является то, что каждый конкретный сервер имен не должен содержать в своей локальной базе данных описание всей иерархии имен DNS. Как правило, организация, регистрирующая свое доменное имя, должна иметь DNS-сервер, в базе данных которого представлено только пространство имен, принадлежащих ее домену (зона ответственности). Например, если в домене «center.fio.ru» имеется только один узел – «www», то в базе данных сервера имен этого домена может присутствовать всего одна запись для указанного имени.

Кроме того, если домен содержит в себе домены нижних уровней, то каждый такой субдомен может иметь свой собственный сервер имен, освобождая тем самым DNS-сервер родительского домена от необходимости обслуживать свое подпространство имен. Таким образом, продолжая пример, сервер имен домена «fio.ru» не имеет в своей базе данных информации об узле «www» домена «center.fio.ru». При этом в его базе должна присутствовать запись о том, что зона «center» обслуживается другим сервером.

Такая передача полномочий по управлению именами части зоны называется *делегированием*.

Система DNS допускает, что один сервер имен может в своей базе данных описывать несколько зон. Например, зоны «center.fio.ru» и «msk.net.fio.ru» обслуживаются одним DNS-сервером.

## 2.4. Разрешение имен

Алгоритм разрешения имен достаточно прост. Когда программному клиенту требуется по доменному имени выяснить IP-адрес, она связывается с сервером имен, адрес которого указан в настройках TCP/IP.

Чтобы программное обеспечение пользовательского компьютера могло осуществлять преобразование доменных имен в IP-адреса, в настройках TCP/IP обязательно должен быть указан адрес хотя бы одного сервера имен.

Сервер имен, получив запрос, рассматривает его, чтобы выяснить, в каком домене находится указанное имя. Если указанный домен входит в его зону ответственности, то сервер преобразует имя в IP-адрес на основе собственной базы данных и возвращает результат клиенту. В случае же, когда сервер имен не способен самостоятельно осуществить преобразование из-за того, что запрашиваемое доменное имя не входит в его зону, он опрашивает известные ему другие сервера имен с целью получения результата.

Для функционирования серверу имен не обязательно знать адреса всех остальных DNS-серверов Интернет. Достаточно располагать адресами серверов имен корневого домена. Как правило, эта информация изначально и постоянно присутствует в программах-серверах. Очевидно также, что сервер имен должен знать адреса DNS-серверов делегированных зон.

Порядок взаимодействия DNS-клиента с сервером для обеспечения разрешения имен определяется специальными протоколами DNS. Этот протокол предусматривает свой формат сообщения (пакета) и использует для доставки данных транспортные протоколы UDP и TCP как нижележащие.

Задачи, решаемые сервисом DNS, являются относительно простыми, поэтому DNS-сообщение устроено несложно и включает в себя:

- поле заголовка, определяющее тип сообщения (например, «запрос клиента», «ответ сервера» и т. д.);
- поле запроса, в котором клиент указывает разрешаемое имя и параметры запроса;
- поле ответа, в которое сервер помещает результат обработки запроса;
- два служебных поля передачи для управляющей и дополнительной информации.

## 2.5. Порядок регистрации доменных имен

Общую координацию регистрации доменных имен выполняет Интернет Корпорация по распределению адресов и имен (The Internet Corporation for Assigned Names and Numbers, ICANN). Эта организация

осуществляет регистрацию имен верхнего уровня. Соответственно с принципами делегирования зон ответственности, принятыми в DNS, за обслуживание пространства имен в каждом домене первого уровня отвечают различные организации. В частности, за зону домена «ru» в настоящее время отвечает Региональный сетевой информационный центр «RU-CENTER». Поэтому, если у пользователя сети Интернет возникла необходимость зарегистрировать доменное имя второго уровня внутри домена «ru», то он должен обращаться именно в эту организацию. Условия и порядок регистрации имен нижних уровней определяют владельцы имен доменов, внутри которых регистрируется новое имя. Например, порядок регистрации имен внутри домена «fio.ru» определяет Федерация Интернет Образования.

При регистрации нового домена независимо от того, где он регистрируется, должны соблюдаться следующие правила:

- имя должно быть уникальным в рамках того домена, где оно регистрируется;
- должен определяться один DNS-сервер (или несколько – для надежности), который будет обслуживать пространство имен создаваемого домена и которому будет делегирована соответствующая зона.

Конкретный порядок задается той организацией, которая осуществляет регистрацию. Так, например, для регистрации домена второго уровня в зоне «ru» через «RU-CENTER» необходимо развернуть два DNS-сервера в различных IP-сетях. При этом указанные сервера должны иметь, по определению Центра, «надежное подключение» к сети Интернет, т. е. связь с серверами не должна прерываться более чем на 2 час в сутки. Если такой возможности нет, например, организация, желающая зарегистрировать имя, не имеет двух сетей с различными номерами, «RU-CENTER» предлагает взять на себя обеспечение функционирования второго сервера имен, а при необходимости – и первого.

### **Методические указания**

Этот раздел пособия необходим для понимания общих принципов построения и функционирования системы DNS. Он содержит следующие ключевые моменты:

- основная цель разработки DNS – обеспечение возможности использования для адресации узлов удобно запоминаемых мнемонических имен вместо числовых IP-адресов;
- основные функции DNS – это организация иерархического пространства имен узлов и обеспечение разрешения имен узлов в IP-адреса;

- принцип иерархической организации основан на том, что узлы сети могут объединяться в логические группы – домены, которые, в свою очередь, могут входить в домены более высокого уровня;
- на самом верхнем уровне иерархии доменов расположен корневой домен, включающий в себя три группы доменов первого уровня: домены общего назначения, национальные домены и обратный домен;
- для адресации конкретного узла используются полные доменные имена, отражающие всю иерархию вхождения узла в домены различных уровней;
- любой узел может входить в несколько доменов и, следовательно, иметь несколько разных полностью определенных доменных имен;
- DNS как сервис разрешения имен строится по принципу «клиент-сервер» и подразумевает наличие в своей структуре клиентского программного обеспечения и серверов имен;
- каждый сервер имен имеет локальную базу данных, в которой описаны соответствия доменных имен и IP-адресов;
- в базе данных сервера имен описываются соответствия не для всего иерархического пространства имен, а для некоторой его части;
- часть пространства имен, обслуживаемая одним сервером называется зоной ответственности;
- алгоритм разрешения предполагает, что клиент направляет запрос серверу, который осуществляет поиск соответствия доменного имени IP-адресу в локальной базе данных;
- в случае отсутствия необходимых сведений в локальной базе данных, сервер имен для подготовки ответа клиенту может связаться с другими серверами.

## ГЛАВА 3. СЕРВИС TELNET – УПРАВЛЕНИЕ УДАЛЕННЫМИ КОМПЬЮТЕРАМИ В ТЕРМИНАЛЬНОМ РЕЖИМЕ

### 3.1. Назначение сервиса Telnet

Сервис *Telnet* – одна из самых старых информационных технологий Internet. В настоящее время этот сервис практически не используется. Поэтому имеет смысл ограничиться изложением только общих принципов организации этого сервиса.

Основным назначением сервиса Telnet является реализация сетевого терминала для доступа к ресурсам удаленного компьютера. Он обеспечивает двунаправленный канал передачи данных (рис. 2).



Рис. 2. Схема клиент-сервер для сервиса Telnet

В сервисе Telnet используется TCP-соединение для передачи данных и управляющей информации. При этом протокол Telnet для обеспечения удаленного доступа терминала к серверу резервирует порт 23. Он выполняет три базовые функции:

1. Определяет сетевой виртуальный терминал (NVT – network virtual terminal), который обеспечивает стандартный интерфейс к удаленной системе.
2. Включает механизм, который позволяет клиенту и серверу согласовать опции обмена.
3. Обеспечивает симметрию соединения, создавая любой программе сервера возможность выступать в качестве клиента.

Протокол Telnet позволяет серверу рассматривать все удаленные терминалы как стандартные «сетевые виртуальные терминалы» строчного типа (рис. 3), работающие в кодах ASCII, а также обеспечивает возможность согласования более сложных функций (например, локальный или удаленный эхо-контроль, страничный режим, и т. д.).



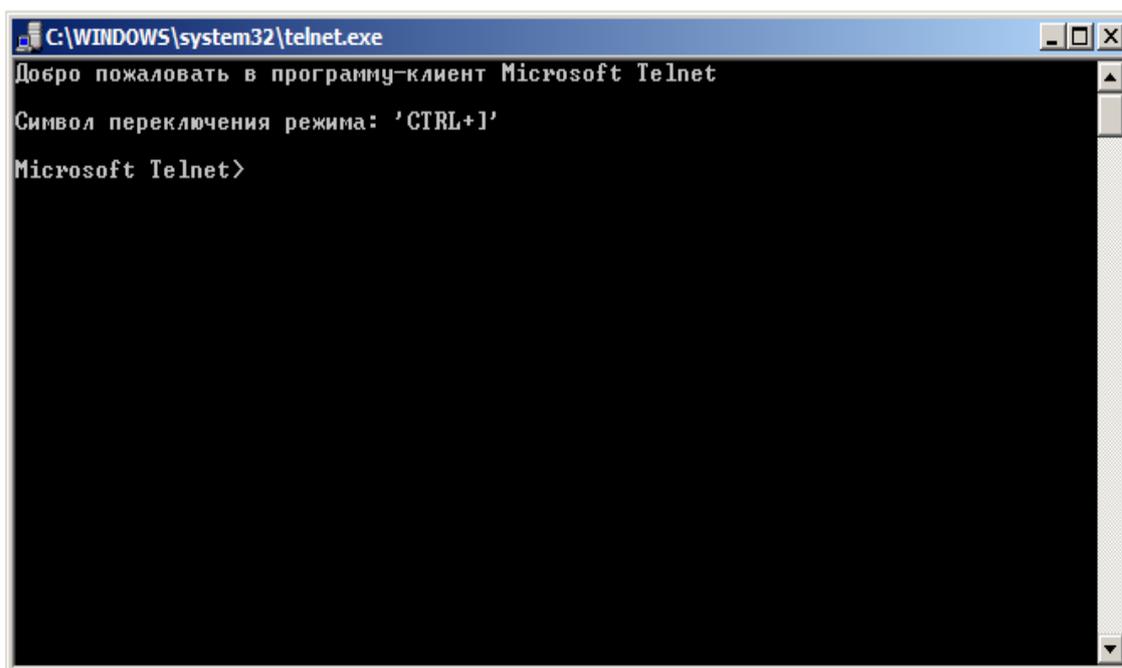
Рис. 3. Схема сетевого виртуального терминала NVT

### 3.2. Организация работы сервиса Telnet

На прикладном уровне над Telnet находится либо программа поддержки реального терминала, либо прикладной процесс на сервере, к которому осуществляется доступ с терминала. Формат NVT достаточно прост. Для данных используются 7-битовые ASCII коды. 8-битовые октеты зарезервированы для командных последовательностей.

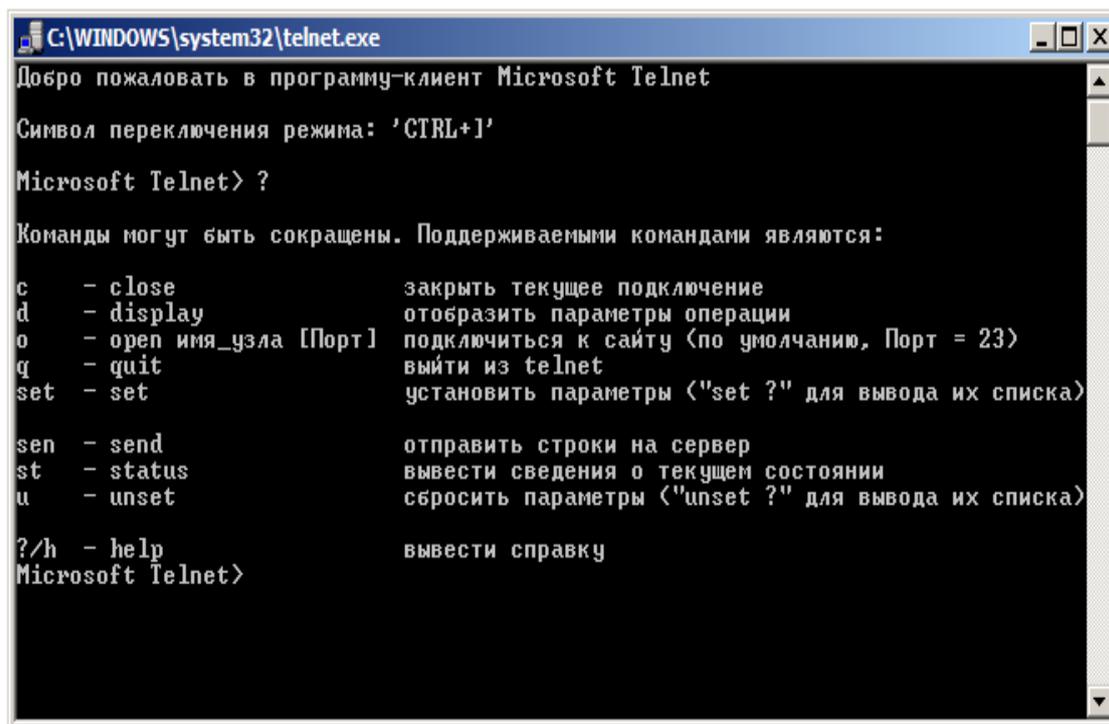
В качестве примера кратко опишем работу программы *telnet*, входящей в поставку ОС Windows. Эта программа предназначена для взаимодействия с другим узлом сети по протоколу Telnet. Чтобы запустить программу, необходимо войти в меню *Пуск* и выбрать пункт *Выполнить*.

Затем вводится в строке ввода имя команды «telnet» и нажимается Enter. Появится консольное окно Windows, в котором выполняется программа telnet.



«Символ переключения режима: 'Ctrl+]'» означает, что для прерывания сеанса работы и возврата к приглашению telnet можно использовать комбинацию клавиш Ctrl+].

Основные команды telnet:



```
C:\WINDOWS\system32\telnet.exe
Добро пожаловать в программу-клиент Microsoft Telnet
Символ переключения режима: 'CTRL+]'
Microsoft Telnet> ?
Команды могут быть сокращены. Поддерживаемыми командами являются:
c      - close           закрыть текущее подключение
d      - display        отобразить параметры операции
o      - open имя_узла [Порт]  подключиться к сайту (по умолчанию, Порт = 23)
q      - quit           выйти из telnet
set    - set            установить параметры ("set ?" для вывода их списка)

sen    - send           отправить строки на сервер
st     - status         вывести сведения о текущем состоянии
u      - unset          сбросить параметры ("unset ?" для вывода их списка)

?/h   - help           вывести справку
Microsoft Telnet>
```

Команды протокола Telnet и принципы работы Telnet-клиентов достаточно подробно изложены в [2, 4].

### Методические указания

В связи с ограниченным использованием сервиса Telnet в данном разделе пособия изложены только общие принципы его организации:

- основным назначением сервиса Telnet является реализация сетевого терминала для доступа к ресурсам удаленного компьютера;
- протокол Telnet выполняет три базовые функции:
  - обеспечивает стандартный интерфейс к удаленной системе посредством сетевого виртуального терминала (NVT);
  - осуществляет согласование опций обмена между клиентом и сервером;
  - создает возможность любой программе сервера выступать в качестве клиента.
- ОС Windows содержит программу telnet, предназначенную для взаимодействия с другими узлами сети по протоколу Telnet.

## ГЛАВА 4. СЕРВИС FTP – СИСТЕМА ФАЙЛОВЫХ АРХИВОВ

### 4.1. Назначение сервиса FTP

Основой сервиса *FTP* является протокол FTP (File Transfer Protocol), который предназначен для передачи файлов между двумя узлами сети. Он предоставляет широкий набор команд для копирования файлов, а также для управления файлами и каталогами на удаленном компьютере: удаление файлов и каталогов, создание каталогов, переименование файлов, получение списка файлов с удаленного компьютера. В случае, если передача файла была прервана по каким-либо причинам, протокол предусматривает средства для докачки файла, что бывает очень удобно при передаче больших файлов.

В предоставлении сервиса FTP принимают участие две программные компоненты:

- *FTP-сервер*. Это программа, которая выполняется на компьютере, предоставляющем сервис FTP. Она принимает команды по протоколу FTP и выполняет их.
- *FTP-клиент*. Это программа, которая выполняется на компьютере пользователя. С ее помощью пользователь устанавливает соединение с *FTP-сервером* и передает ему различные команды: запрашивает файлы для скачивания, удаляет файлы, переименовывает их и т. д.

Общая схема работы сервиса FTP приведена на рис. 4. FTP-сервер ожидает соединения от клиента на определенном порту (обычно используется порт 21, хотя это не обязательно). Когда соединение установлено, клиент использует его для передачи команд серверу, а сервер для передачи ответов клиенту.



Рис. 4. Схема работы сервиса FTP

Это так называемое «управляющее» TCP соединение. В отличие от других протоколов, таких как SMTP или HTTP, для передачи данных

используется отдельное, второе TCP соединение. Данные по управляющему соединению передаются в соответствии с протоколом Telnet.

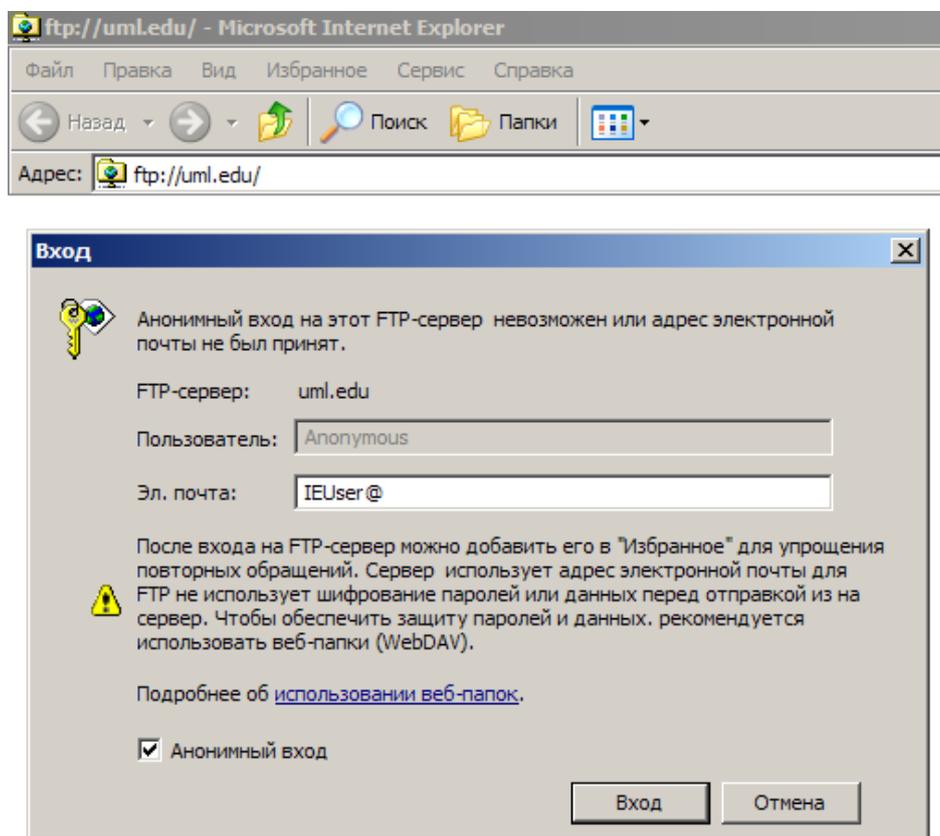
Установив управляющее соединение с сервером, клиент должен провести процедуру авторизации: сообщить имя своей учетной записи и пароль.

## 4.2. Организация работы сервиса FTP

Работа клиента с FTP-сервером может осуществляться как в командном режиме (аналогично сервису Telnet), так и в режиме FTP-клиента, встроенного в Internet Explorer. Команды протокола FTP и принципы работы FTP-клиентов в командном режиме достаточно подробно изложены в [2, 4]. В рамках данного пособия имеет смысл ограничиться изложением только второго режима. В этом режиме работа с FTP-архивами практически не отличается от работы с файлами на компьютере клиента.

В начале работы открывается окно Internet Explorer и устанавливается соединение с FTP-сервером. Для этого в строке адреса вводится: ftp://имя-FTP-сервера/. После установки соединения Internet Explorer запросит пароль, соответствующий указанной учетной записи.

Существует множество так называемых «анонимных» FTP-серверов, которые разрешают только скачивать файлы на свой компьютер, но не позволяют загружать файлы на сервер. При регистрации на таких серверах используется имя учетной записи «anonymous» или «ftp».

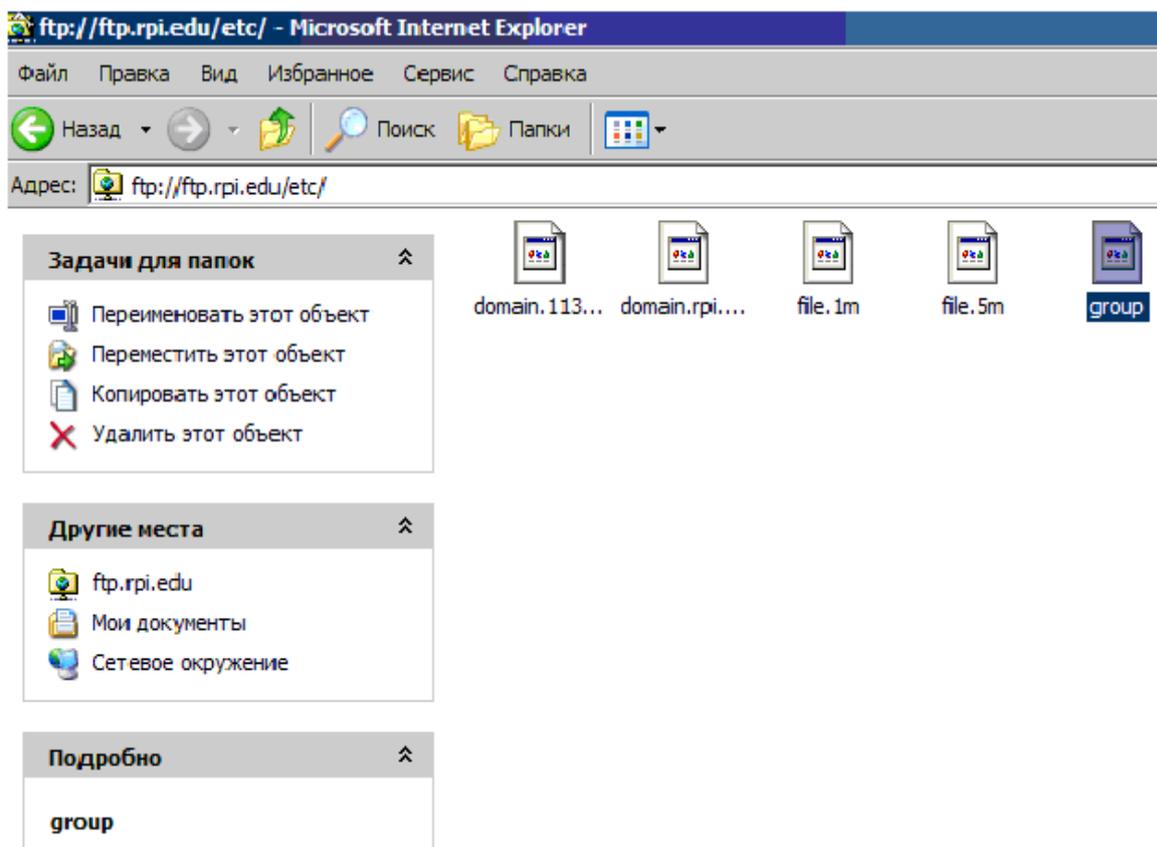


Ссылка на объект, доступный через анонимное FTP, обычно записывается в следующем виде (табл. 3):

Таблица 3

*Пример обращения к ресурсу на FTP-сервере*

Имя сервера	Имя каталога в сервере
ftp://ftp.rpi.edu/	/etc/group/



Если соединение установлено успешно, то можно скачивать файлы на свой компьютер и загружать их на сервер точно так же, как при копировании на локальном диске клиента.

### Методические указания

Как и сервис Telnet, сервис FTP является старой информационной технологией Internet. Поэтому в данном разделе пособия приведены только общие понятия этого сервиса:

- основой сервиса FTP является протокол FTP, который предназначен для передачи файлов между двумя узлами сети;
- в предоставлении сервиса FTP принимают участие две программные компоненты: FTP-сервер и FTP-клиент;

- работа клиента с FTP-сервером может осуществляться как в командном режиме (аналогично сервису Telnet), так и в режиме FTP-клиента, встроенного в Internet Explorer;
- существует множество так называемых «анонимных» FTP-серверов, которые разрешают скачивать файлы на свой компьютер без указания пароля.

## ГЛАВА 5. СЕРВИС USENET – СИСТЕМА ТЕЛЕКОНФЕРЕНЦИЙ

### 5.1. Назначение сервиса Usenet

*Usenet* – это система телеконференций (или группы новостей), представляющая собой распределенную информационную систему. В Usenet нет центрального места хранения, и информация осуществляет постоянное движение в сети от сервера к серверу, от сервера к клиенту, от клиента к серверу. Когда пользователь отправляет сообщение, вначале оно становится доступно только на его сервере. Но каждый сервер *news* обменивается сообщениями с несколькими соседними серверами, и таким образом сообщение оказывается на каждом сервере в сети.

В настоящее время практически весь Usenet-трафик передаётся по Интернету, а формат сообщений и способ их передачи очень похож на электронную почту. Однако, если электронная почта используется прежде всего для общения «один на один», то Usenet действует по принципу «один для всех».

Сообщения, которые пользователь публикует в Usenet, организуются в тематические категории, называемые конференциями или новостными группами. Сообщения в группе обычно не задерживаются более нескольких дней (стандартное значение по умолчанию – 5).

Группы новостей образуют собственную иерархию, подобную структуре доменных имён. Например, группы *sci.math* и *sci.physics* находятся внутри иерархии *sci* (сокращение от англ. *science* – наука).

Основные иерархии конференций Usenet составляют «Большую Восьмёрку»:

- *comp.\**: обсуждение тем, связанных с компьютерами (*comp.software*, *comp.sys.amiga*);
- *misc.\**: разнообразные темы (*misc.education*, *misc.forsale*, *misc.kids*);
- *news.\**: новости Usenet-a (*news.groups*, *news.admin*);
- *rec.\**: развлечения и отдых (*rec.music*, *rec.arts.movies*);
- *sci.\**: научные дискуссии (*sci.psychology*, *sci.research*);
- *soc.\**: социальные темы (*soc.college.org*, *soc.culture.african*);
- *talk.\**: разговоры, в том числе на «горячие» темы (*talk.religion*, *talk.politics*);
- *humanities.\**: искусство, литература, философия (*humanities.classics*, *humanities.design.misc*).

С помощью приложений для работы с Usenet пользователь может «подписаться» на любые доступные конференции. Он может выбрать

одну или несколько подгрупп и осуществить на них подписку. Подписка подразумевает процедуру оповещения пользователя о появлении новых статей по интересующей его теме. Естественно, что пользователь и сам может отправить свое мнение, оформленное в виде сообщения, в группу. Сообщение оформляется в соответствии со стандартом почтового сообщения Internet. Фактически Usenet выполняет функции периодического издания.

## 5.2. Организация подписки на конференции

В настоящее время существует несколько тысяч групп – от достаточно серьезных и профессионально-ориентированных до групп посвященных играм, шуткам и т. д.

### *Список некоторых серверов*

#### *Платные*

- Usenext <http://www.usenext.com/> – ~80 дней
- AlphaLoad <http://alphaload.de/> – ~30 дней
- firstload <http://firstload.de/> – ~30 дней
- aEton <http://aeton.de/> – ~30 дней
- EasyUsenet <http://www.easyusenet.nl/> – 100 дней
- Usenet.com <http://www.usenet.com/> – 75 дней
- PowerUsenet <http://www.powerusenet.com/> – 75 дней
- NewsHosting <http://www.newshosting.com/> – 70 дней
- ngroups.net <http://ngroups.net/> – 75 дней
- newsfeeds.com <http://www.newsfeeds.com/> – 75 дней
- UseNetServer <http://www.usenetserver.com/> – 101 день
- Alibis <http://www.alibis.com/> – 75 дней
- Usenet-News <http://usenet-news.net/> – 75 дней
- Binaries.net <http://www.binaries.net/> – ~30 дней
- eurousenet <http://www.eurousenet.com/> – 55 дней
- Meganetnews <http://www.meganetnews.com/> – ~30 дней
- Newsgroups.com <http://www.newsgroups.com/> – 31 день
- NewsGuy <http://www.newsguy.com/> – 90 дней
- TweakNews <http://tweaknews.nl/> – 50 дней
- UsenetNewsServer <http://www.usenetnewsserver.com/> – 90 дней
- Astraweb <http://news.astraweb.com/> – 68 дней
- NewsDemon <http://www.newsdemon.com/> – 110 дней
- Teranews <http://www.teranews.com/> – ~30 дней

## Бесплатные

- <http://conf.msu.ru> – научные конференции в России и других странах.
- Google Groups (<http://groups.google.com/>) – позволяет работать только с текстовыми группами с помощью web-интерфейса.
- <http://www.sforum.nl/nntp/show.php?l=en>
- <http://www.binaryfeeds.com/>

Для подписки на конференцию (например, научные конференции в России и других странах) необходимо обратиться на бесплатный сервер <http://conf.msu.ru> с помощью Internet Explorer.

### "КОНФЕРЕНЦИЯ"

**все о научных конференциях в России и в мире**

[В начало](#) [Конференции](#) [Организатору](#) [Настройка](#) [Подписка](#)

>> Начало:

**Основные разделы**

- [Кибернетика \(4\)](#)
- [Математика \(3\)](#)
- [Информатика \(3\)](#)
- [Автоматика, вычислительная техника \(2\)](#)
- [Народное образование, педагогика \(2\)](#)
- [Экономика, экономические науки \(2\)](#)
- [Статистика \(1\)](#)
- [Организация и управление \(1\)](#)
- [Общие и комплексные проблемы технических и прикладных наук и отраслей народного хозяйства \(1\)](#)
- [Медицина и здравоохранение \(1\)](#)
- [Приборостроение \(1\)](#)
- [Связь \(1\)](#)

**Последние поступления**

- [Конкуренция и конкурентоспособность. Организация производства конкурентоспособной продукции.](#)
- Проблемы геологии, экологии и рационального природопользования
- [Современная техника и технологии в медицине, биологии и экологии](#)
- [Компьютерные технологии в науке, производстве, социальных и экономических процессах](#)
- МАТЕМАТИКА. КОМПЬЮТЕР. ОБРАЗОВАНИЕ**
- [Естественнонаучное образование в средней школе](#)
- [Проблемы синергетики в трибологии, трибозлектрохимии, материаловедении и мехатронике](#)
- [Микропроцессорные, аналоговые и цифровые системы: проектирование и схемотехника, теория и вопросы применения](#)
- [Интеллектуальные электромеханические устройства, системы и комплексы](#)
- Теория, методы и средства измерений, контроля и диагностики

**Поиск** прошедшие

**Новости**

[19.02.2004] Добавлен поиск по тематике конференций.

**Статистика**

Всего конференций	1368
Зарегистрировано пользователей	6673
Количество посещений	310978
Показано конференций	670479

Если у Вас есть информация о конференциях, [присылайте](#) ее для размещения на сервере.

Сервер разработан и поддерживается при содействии Российского Фонда Фундаментальных Исследований (грант N00-07-90185-в)

Все вопросы и пожелания просьба направлять [администратору](#) системы "Конференция"

Copyright ©1999-2004

Затем нажать на гипертекстовую ссылку Подписка. Появится окно для регистрации пользователя.

## "КОНФЕРЕНЦИЯ"

**все о научных конференциях в России и в мире**

[В начало Конференции](#) [Организатору](#) [Настройка](#) [Подписка](#)

>> Регистрация:

**Вход для зарегистрированных пользователей.**

Если Вы еще не зарегистрировались, заполните форму. Подтверждение о регистрации и пароль будет выслан по указанному Вами адресу E-mail.

**Введите данные для регистрации**

Фино (\*)

Идентификатор (\*) E-mail (\*)

Пароль (\*)      Повторите пароль (\*)

Copyright ©1999-2004

После регистрации пользователь получает подтверждение о регистрации.

[В начало Конференции](#) [Организатору](#) [Настройка](#) [Подписка](#)

>> Регистрация:

Регистрация прошла успешно!

В ближайшее время по указанному Вами адресу электронной почты ([kvp@osu.cctpu.edu.ru](mailto:kvp@osu.cctpu.edu.ru)) Вам будет выслано письмо с указанием ссылки, которую Вам необходимо посетить для активации режима подписки.

Copyright ©1999-2004

### 5.3. Организация работы сервиса Usenet

Загрузка и публикация файлов в Usenet выполняется следующим образом. Пользователь осуществляет подписку на одном из серверов Usenet, который ближе для данного пользователя (рис. 5). Обычно это компьютер, на котором расположены все информационные ресурсы организации или учебного заведения. По мере поступления новых сообщений от пользователей, серверы обмениваются между собой этой но-

вой информацией. Пользователь может подписаться на любом сервере на любую группу новостей. Каждый сервер ведет список подписчиков и список телеконференций, на которые осуществлена подписка. При поступлении статьи, сервер информирует известные ему серверы о том, что появилась новая информация и в случае запроса передает ее.

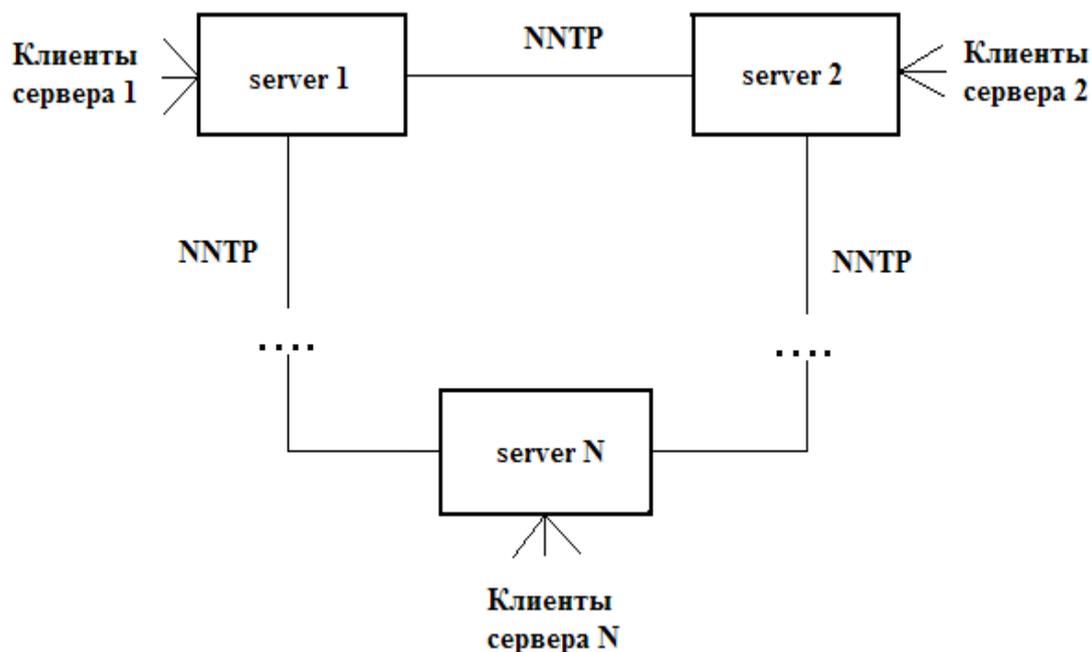


Рис. 5. Схема организации серверов Usenet

Сервера Usenet дублируют (частично или полностью) информацию хранящуюся друг у друга для того чтобы их пользователи могли общаться между собой. Однако, работают они не синхронно. Каждый из них имеет свои ограничения на объём загружаемой ежедневно информации с других серверов или, по крайней мере, сроки её хранения (*retention rate*). Причем администрация сервера определяет сама, какие группы следует синхронизировать, в каком объёме и сколько в той или иной группе должна храниться информация. Каждый сервис в силу собственных возможностей стремится предоставить максимум своим пользователям. Многие платные серверы негласно приветствуют практически любой контент, даже пиратский, отказываясь вести логи (протоколировать действия своих пользователей) и не имеют ограничений на публикацию файлов.

Взаимодействие серверов осуществляется посредством протокола NNTP. Протокол NNTP пришел на смену UUCP, его целью было упорядочить обмен информацией между серверами Usenet. Кроме этого, использование интерактивного протокола позволило разработать интерфейсные программы-клиенты, позволяющие просматривать почту на удаленных серверах.

При передаче новостей по протоколу UUCP все новые сообщения передавались от одного сервера другому, затем этого сервера следующему и т. д. В результате в сети возникало столько одинаковых баз данных новостей, сколько имелось в наличии серверов. При этом на сервер попадала информация, которая либо никому на этом сервере не была нужна, либо уже имелась. Понятно, что сеть в этом случае загружалась непроизводительно.

С использованием NNTP ситуация существенно изменилась. При использовании интерактивного протокола программы-серверы могут обмениваться информацией о наличии сообщений и заказывать только те, на которые есть запросы. В протоколе вводятся два уровня серверов. Центральный сервер обеспечивает централизованное хранение новостей в локальных сетях. Пользователи используют программы-клиенты для просмотра новостей. На этом же сервере хранятся и списки подписки пользователей на группы новостей. Промежуточный или «slave» (подчиненный) сервер обычно устанавливается на машинах с большим числом пользователей и хранит только последние поступления. За всей остальной информацией он обращается к центральному серверу сети.

### **Методические указания**

В этом разделе пособия рассмотрены вопросы назначения и подписки на телеконференции, а также организации их функционирования. При изучении этого раздела необходимо усвоить следующее:

- Usenet – это система телеконференций (или группы новостей), в которой каждый сервер *news* обменивается сообщениями с несколькими соседними серверами, и таким образом, сообщение оказывается на каждом сервере в сети;
- сообщения, которые пользователь публикует в Usenet, организуются в тематические категории, называемые конференциями или новостными группами;
- группы новостей образуют собственную иерархию, подобную структуре доменных имён;
- сообщения в группе задерживаются только на несколько дней (стандартное значение по умолчанию – 5);
- пользователь может выбрать одну или несколько подгрупп и осуществить на них «подписку», которая подразумевает процедуру оповещения пользователя о появлении новых статей по интересующей его теме;
- взаимодействие серверов Usenet осуществляется посредством протокола NNTP.

## ГЛАВА 6. СЕРВИС E-MAIL – ЭЛЕКТРОННАЯ ПОЧТА

### 6.1. Назначение сервиса E-mail

Сервис *E-mail* предназначен для передачи почтовых сообщений в электронном виде. Организация сообщения E-mail аналогична структуре обычного почтового сообщения. Оно состоит из заголовка («конверта»), содержащего информацию, обеспечивающую доставку (например, адреса получателя и отправителя), и тела сообщения – блока информации, который необходимо передать получателю. В простейшем случае тело сообщения содержит просто текст. Однако при необходимости отправитель может «прикрепить» дополнительные данные, представленные в виде дисковых файлов. Такие данные называются вложениями (attachments).

Сервис E-mail включает в себя следующие основные компоненты:

1. *Почтовый сервер* – программный комплекс, обеспечивающий пересылку сообщений электронной почты, их хранение, управление базой данных сообщений, а также позволяющий абоненту работать со своим почтовым ящиком. В настоящее время функционирование электронной почты реализуется с помощью нескольких протоколов, поэтому в составе почтовых серверов обычно присутствуют несколько модулей – серверных частей, реализующих взаимодействие по основным почтовым протоколам: SMTP-сервер, выполняющий функции сервера исходящей почты, а также POP3-сервер и/или IMAP4-сервер, представляющие собой серверы входящей почты.
2. *Почтовый клиент* – программа, позволяющая пользователю создавать и отправлять сообщения, а также управлять полученными сообщениями (просматривать, упорядочивать, печатать, удалять и т. д.) из собственного почтового ящика. Часто программный модуль, обеспечивающий взаимодействие с почтовым сервером, называют пользовательским агентом (User Agent, UA). Современные почтовые клиенты позволяют взаимодействовать с серверными модулями основных протоколов электронной почты (SMTP, POP3 и IMAP4).
3. *Протоколы взаимодействия почтовых клиентов с серверами.* В настоящее время основными протоколами являются:
  - *SMTP (Simple Mail Transfer Protocol)*, использующийся сервисом электронной почты для передачи сообщений от отправителя к получателю.
  - *POP3 (Post Office Protocol version 3)* и *IMAP4 (Internet Message Access Protocol version 4)*, обеспечивающих выборку входящих сообщений из почтового ящика. Протокол POP3 является более

простым, менее функциональным и защищенным, нежели IMAP4. Однако IMAP4 в настоящее время поддерживается не всеми почтовыми серверами и клиентами и поэтому является менее популярным.

Схема взаимодействия основных компонент E-mail приведена на рис. 6.

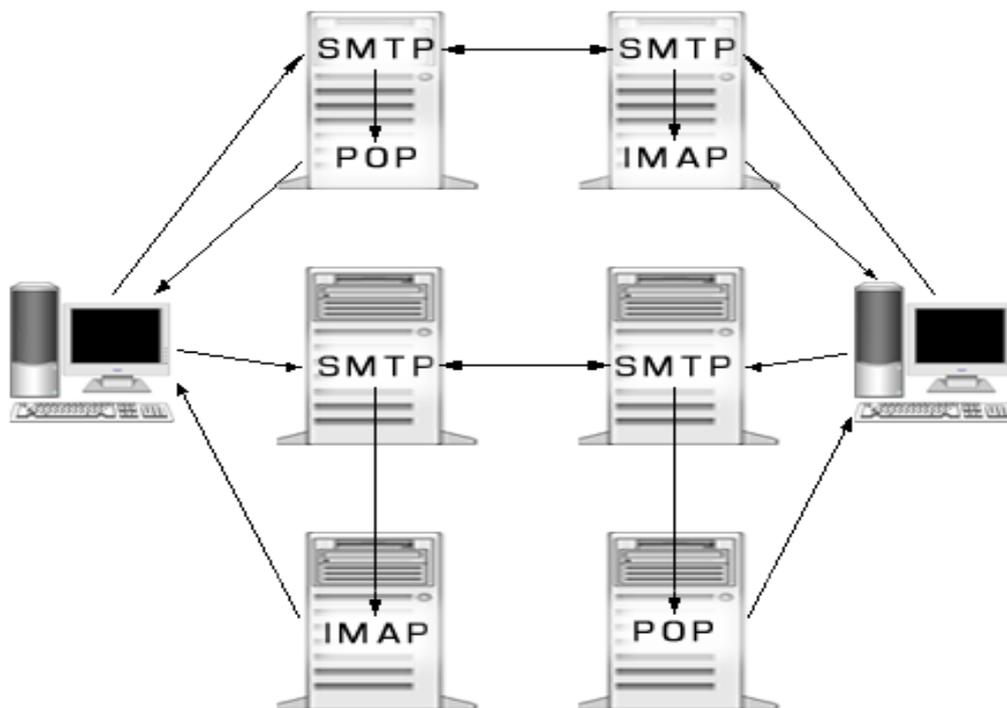


Рис. 6. Схема взаимодействия основных компонент E-mail

## 6.2. Стандартные функции почтовых клиентов

### 6.2.1. Прием сообщений

Существуют три базовых модели, описывающих взаимодействие почтового клиента и сервера: *автономная, интерактивная и отключенная*:

- *Автономная (offline) модель* предполагает, что клиент периодически подключается к серверу для получения почты. После выборки клиентом сообщения обрабатываются на узле получателя и могут быть удалены с сервера.

С одной стороны, удаление сообщений может оказаться полезным, т. к. объем почтового ящика, как правило, ограничен, и накопление в нем сообщений может вызвать переполнение. И как следствие – невозможность получения новых сообщений. С другой стороны, удаление писем с сервера после выборки их клиентом делает невозможной работу с ними с другого узла сети. Поэтому современные почтовые клиенты,

взаимодействующие с сервером по этой модели, предоставляют пользователю возможность установить или отключить режим удаления сообщений после выборки.

- *Интерактивная (online) модель* основана на том, что вся почта хранится и обрабатывается на сервере, а клиент подключается и во время сеанса управляет этим процессом. Поскольку хранение сообщений непосредственно в почтовом ящике может привести к его быстрому переполнению, то пользователю необходимо следить за этим и своевременно удалять ненужные письма.
- *Отключенная (disconnected) модель* представляет собой нечто среднее между автономной и интерактивной моделями. Она предполагает, что клиент подключается к серверу, принимает выборочно сообщения и обрабатывает их в автономном режиме. После этого клиент может снова подключиться к серверу и передать ему изменения. Для обеспечения этой функции в состав почтового клиента должны быть включены программные модули клиентских частей основных протоколов получения почты: POP3 и/или IMAP4.

Протокол POP3 позволяет организовать взаимодействие по автономной модели, а IMAP4 – по всем трем моделям. Однако на практике почтовые клиенты, взаимодействующие с сервером по протоколу IMAP4 в соответствии с принципами автономной модели, не встречаются, поскольку такое решение является неэффективным.

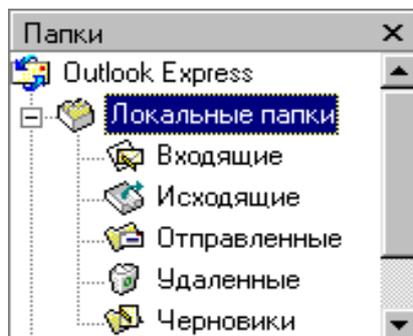
### **6.2.2. Локальное хранение сообщений**

Для того, чтобы пользователь имел возможность работать с сообщениями в автономном режиме (просматривать полученные, создавать и редактировать новые), почтовый клиент должен обеспечивать хранение почты на пользовательском узле. Для удобства работы пользователя клиентская программа должна также определенным образом упорядочивать сообщения. Для этого локальное хранилище должно предусматривать возможность создания папок. Как правило, почтовые клиенты предлагают стандартный набор папок (рис. 7):

- папка «Входящие» («Inbox») – предназначена для хранения вновь поступившей почты;
- папка «Исходящие» («Outbox», «Unsent Messages») – используется для хранения сообщений, предназначенных для отправки;
- папка «Отправленные» («Sent», «Sent items», «Sent mail») – в нее помещаются копии отправленных сообщений;
- папка «Удаленные» («Trash», «Deleted items») – в папку перемещаются сообщения, удаленные из других папок, что позволяет восстанавливать ошибочно удаленные данные;

- папка «Черновики» («Drafts») – предназначена для хранения черновых вариантов сообщений, которые пользователь предполагает дорабатывать перед отправкой.

Помимо операций со стандартными папками почтовый клиент должен обеспечивать функции работы с дополнительными папками (создание, удаление, перемещение и т. п.);



*Рис. 7. Папки локального хранилища почтового клиента Microsoft Outlook Express*

### **6.2.3. Создание и отправка сообщений**

Создание сообщений предполагает две группы действий:

- создание заголовка и заполнение основных полей;
- создание собственно сообщения.

Для создания сообщений почтовые клиенты имеют в своем составе специальный редактор, упрощающий ввод основных полей заголовка (часть полей заголовка заполняются автоматически почтовым клиентом), а также позволяющий быстро сформировать текст сообщения (рис. 8).

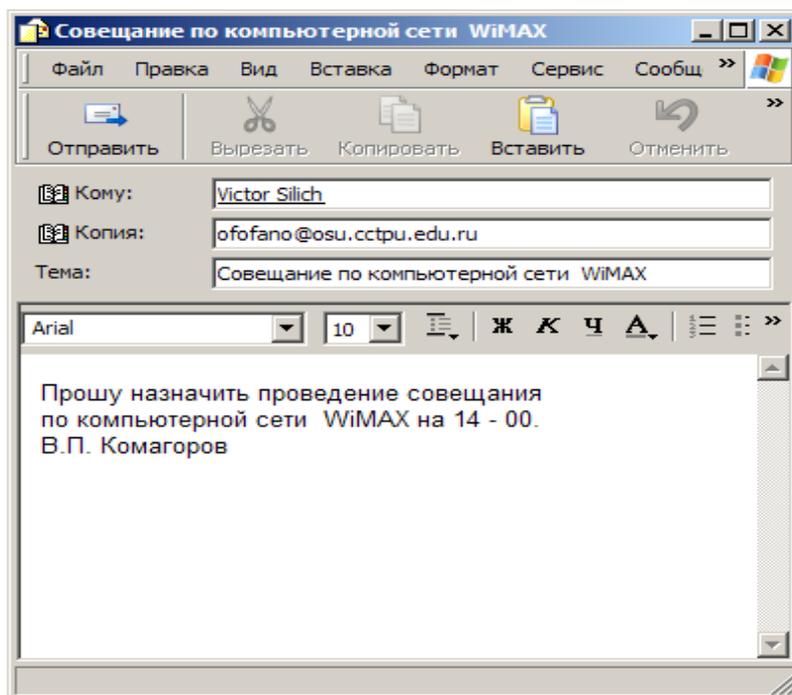
Функциональные возможности редакторов различны и зависят от свойств конкретной программы-клиента. Однако существует ряд требований общего плана, которые должны выдерживаться:

- поддержка выбора кодировки текста;
- поддержка работы с буфером обмена;
- поддержка (как дополнительное свойство) возможности создания форматированного текста, а также прием форматированных сообщений (например, в формате HTML).

Для отправки сообщений в состав почтового клиента должен включаться программный модуль клиентской части протокола SMTP.

Обычно отправка сообщений почтовым клиентом организуется в два этапа:

1. Созданное сообщение по команде пользователя, предназначенной для отправки сообщений, помещается в папку «Исходящие» («Outbox», «Unsent Messages» и т. п.).



*Рис. 8. Окно редактора сообщений*

2. Во время подключения к серверу исходящей почты все сообщения, находящиеся в папке «Исходящие», отправляются на сервер. При этом почтовый клиент можно настроить таким образом, чтобы копия отправленного сообщения сохранялась в папке «Отправленные» («Sent», «Sent items», «Sent mail» и т. п.).

#### **6.2.4. Работа с почтовыми вложениями**

Поддержка работы с почтовыми вложениями обеспечивает пользователю возможность пересылать данные, хранящиеся в отдельных разнотипных файлах. Для этого почтовые клиенты должны иметь в своем составе команды «прикрепления» файла к сообщению и исключения файла из сообщения (при отправке), а также команды, позволяющие сохранить вложение в виде файла на жестком диске (при получении). Часть почтовых клиентов может просмотреть вложенные файлы полученного сообщения без сохранения на диске (рис. 9).

#### **6.2.5. Создание ответных сообщений**

Для упрощения создания ответного сообщения на основе поступившего сообщения почтовые клиенты располагают функцией, автоматизирующей этот процесс следующим образом (рис. 10):

- формируется содержимое тех полей заголовка, которые заполняются автоматически при обычном создании нового сообщения, например дата отправки, адрес отправителя и т. п.;

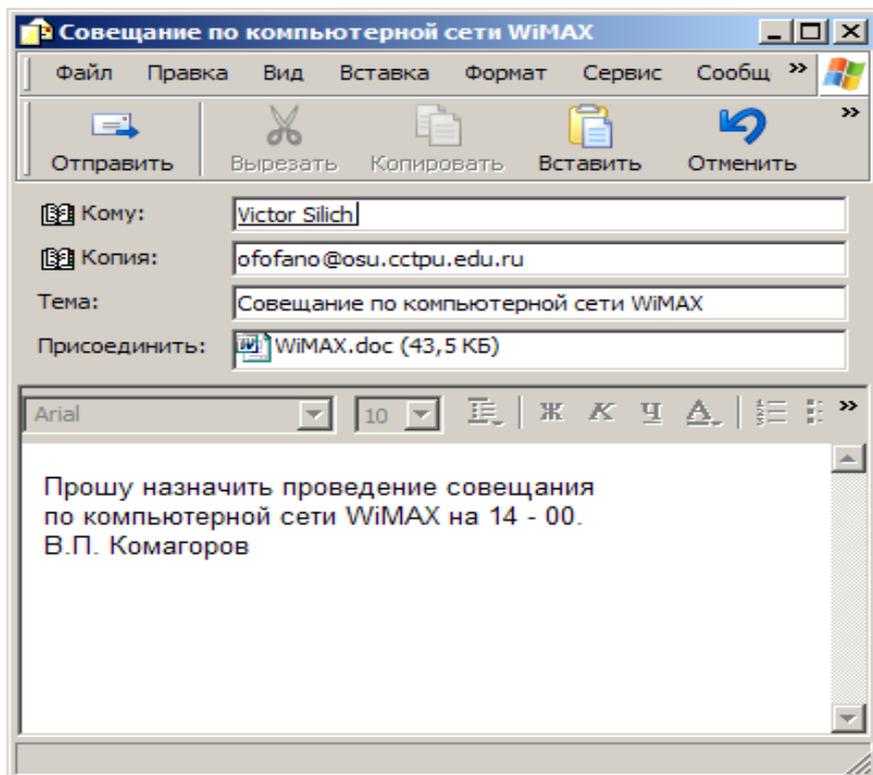


Рис. 9. Окно редактора сообщений с вложениями

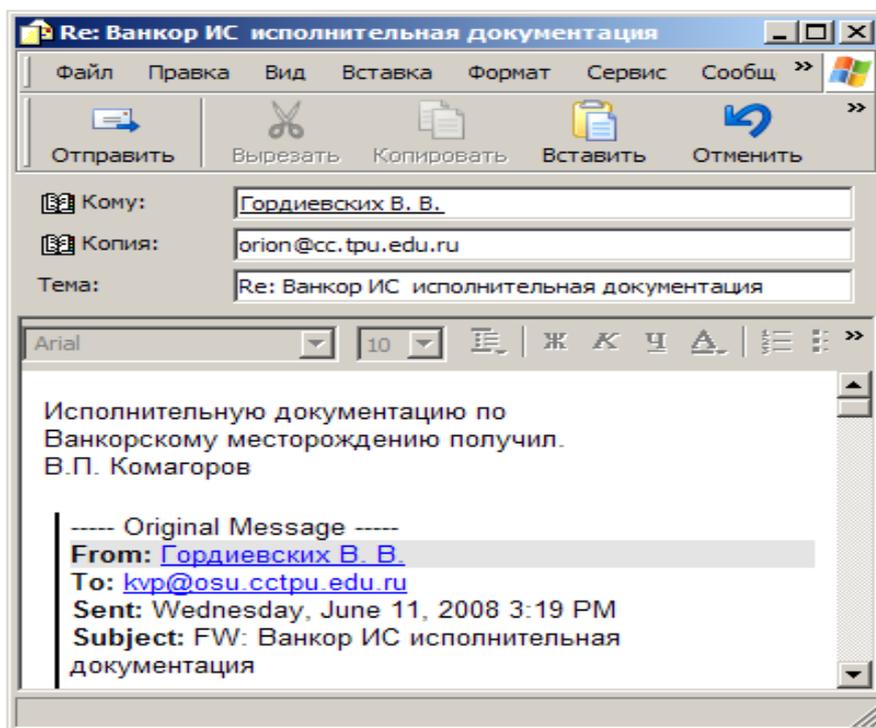


Рис. 10. Окно редактора со сформированным ответным сообщением

- заполняется поле «То» («Кому») ответного сообщения на основе ответного адреса, указанного в поле «Reply-to» («Ответить») заго-

- ловка исходного сообщения, а в случае, если это поле отсутствует – на основе адреса отправителя, указанного в поле «From» («От»);
- формируется содержимое поля «Subject» («Тема») на основе темы исходного сообщения, как правило, по такой схеме: тема исходного сообщения предваряется префиксом «Re:» (от английского слова reply – ответить);
  - в тело сообщения помещается сокращенный вариант заголовка и текст исходного сообщения, каким-либо образом помеченный, чтобы его можно было отличить от текста, который будет вводить отвечающий;
  - исходное сообщение в ответное сообщение не помещаются.

### **Методические указания**

В процессе изучения этого раздела пособия необходимо уяснить следующие моменты:

- сервис E-mail предназначен для передачи почтовых сообщений в электронном виде и включает в себя следующие основные компоненты: почтовый сервер, почтовый клиент и протоколы взаимодействия почтовых клиентов с серверами;
- почтовый сервер – программный комплекс, обеспечивающий пересылку сообщений электронной почты, их хранение, управление базой данных сообщений;
- почтовый клиент – программа, позволяющая пользователю создавать и отправлять сообщения, а также управлять полученными сообщениями (просматривать, упорядочивать, печатать, удалять и т. д.);
- основными протоколами взаимодействия почтовых клиентов с серверами являются протокол SMTP, предназначенный для передачи сообщений от отправителя к получателю, и протоколы POP3 и IMAP4, обеспечивающие выборку входящих сообщений из почтового ящика;
- почтовые клиенты включают в себя следующие стандартные функции: прием сообщений, локальное хранение сообщений, создание и отправка сообщений, работа с почтовыми вложениями, автоматизированное создание ответных сообщений;
- существуют три базовых модели взаимодействия почтового клиента и сервера: автономная, интерактивная и отключенная;
- автономная (offline) модель предполагает, что клиент периодически подключается к серверу и осуществляет выборку сообщений, которые обрабатываются на узле получателя и могут быть удалены с сервера;
- интерактивная (online) модель основана на том, что вся почта хранится и обрабатывается на сервере, а клиент подключается и во время сеанса связи управляет этим процессом;

- отключенная (disconnected) модель предполагает, что клиент подключается к серверу, принимает выборочно сообщения и обрабатывает их в автономном режиме;
- почтовые клиенты предлагают следующий стандартный набор папок для локального хранения сообщений: «Входящие», «Исходящие», «Отправленные», «Удаленные»; «Черновики»;
- для создания сообщений почтовые клиенты имеют в своем составе специальный редактор, позволяющий сформировать текст сообщения;
- отправка сообщений почтовым клиентом обеспечивается с помощью программного модуля клиентской части протокола SMTP;
- почтовые клиенты для работы с почтовыми вложениями имеют в своем составе команды «прикрепления» файла к сообщению;
- создание ответного сообщения осуществляется на основе поступившего сообщения с помощью функции, автоматизирующей этот процесс.

# ГЛАВА 7. СЕРВИС WWW – ГИПЕРТЕКСТОВАЯ СИСТЕМА ИНТЕГРАЦИИ СЕТЕВЫХ РЕСУРСОВ В ЕДИНОЕ ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО

## 7.1. Назначение сервиса WWW

Сервис *World Wide Web* (WWW, W3, Web, «всемирная паутина») является, пожалуй, одним из самых популярных в настоящее время сервисов Интернет. Это можно объяснить тем, что основной концепцией, заложенной при его создании, является концепция «универсальной читаемости» (Universal readership). К 1989 году, когда впервые была предложена идея этой системы, в Интернет уже существовало достаточное количество разнообразных сервисов – электронная почта, FTP, телеконференции, telnet и т. д., каждый из которых позволял манипулировать своими информационными ресурсами. Для того, чтобы иметь возможность работать с разными ресурсами, пользователь был вынужден запускать параллельно, или что еще менее удобно – последовательно, несколько различных клиентских программ. Естественно, что работа в таком режиме оказывалась малоэффективной.

При создании WWW было предложено разработать универсальный клиент – программу, которая могла бы взаимодействовать с серверами различных сервисов – FTP-серверами, NNTP-серверами (телеконференции) и т. д. и предоставлять разнородную информацию пользователю в удобном виде.

Наряду с этим появилась идея создания на основе ресурсов различных сервисов интегрированного информационного пространства. Интеграция ресурсов предусматривала установление взаимосвязей между различными информационными объектами (файлами, сообщениями и т. п.), т. е. по сути – создание своеобразной логической «сети» объектов. Но поскольку объекты, относящиеся к ресурсам существовавших тогда сервисов, не позволяли устанавливать между собой логические связи, то реализация такой идеи потребовала разработки нового способа представления информации, который бы позволил создавать новые информационные объекты (документы) и связывать их с уже существующими. Наиболее удобным для этого представлением данных оказался *гипертекст*.

Классическим примером *гипертекста* являются энциклопедические словари: в тексте статьи предусмотрен механизм «отсылки» читателя к другим статьям, связанных с ней по смыслу. Этот механизм реализуется выделением терминов в тексте статьи. Таким образом, устанавливаются

связи между различными информационными фрагментами. В компьютерных технологиях под гипертекстом понимают форму представления информационных объектов, позволяющую устанавливать связи между фрагментами этих объектов. Механизм, обеспечивающий такую связь, принято называть *гиперссылкой*.

Практическая реализация такого подхода потребовала разработки специального языка для описания, позволяющего описывать внешнее представление гипертекстовых документов и устанавливать связи между информационными объектами. Для обеспечения работы в гипертекстовой среде оказалось необходимым также создание дополнительного программного обеспечения (клиента и сервера) и протокола взаимодействия.

Итак, сервис WWW включает в себя следующие компоненты:

1. *Web-документ (Web-страница)* – гипертекстовый документ, содержащий в себе гиперссылки на другие Web-документы, различные информационные объекты (например, графические, звуковые файлы и т. п.) и ресурсы других сервисов. Адресация объектов обеспечивается с помощью унифицированных указателей ресурса *URL (Uniform Resource Locator)*. Для описания внешнего вида Web-документов и его связей с другими документами и объектами был разработан специальный язык разметки гипертекста *HTML (Hyper Text Markup Language)*.
2. *Web-сайт* – под Web-сайтом понимается совокупность объединенных по смыслу и связанных с помощью гиперссылок Web-документов, обладающих следующими свойствами:
  - целостностью и логической законченностью представления информации;
  - наличием собственного адреса в сети.Поскольку Web-документы позволяют связать не только Web-документы, но и другие информационные ресурсы, то информационное пространство WWW является интегрирующим пространством.
3. *Web-сервер* – серверное программное обеспечение, предназначенное для управления и пересылки по запросу Web-документов клиентам. Основным компонентом Web-сервера является HTTP-сервер – программный сервер, обеспечивающий прием запросов от клиента по протоколу HTTP и формирование ему ответа. Помимо HTTP-сервера в состав Web-сервера могут входить различные программные расширения, например, обеспечивающие динамическое формирование Web-документов.
4. *Web-клиент (браузер)* – клиентская программа, позволяющая получить Web-документ или другой объект, представить его пользователю и обеспечивающая возможность работы с ним. Универсаль-

ность Web-клиента состоит в том, что он «умеет» связываться не только с Web-сервером, но и серверами других сервисов, например, FTP-серверами.

5. *Протокол HTTP (Hyper Text Transfer Protocol)* – протокол взаимодействия клиента с Web-сервером.

## 7.2. Средства представления Web-документов

### 7.2.1. Язык HTML

Для разметки гипертекстовых документов был разработан специальный язык *HTML (Hyper Text Markup Language)*. С момента рождения WWW HTML постоянно развивается. В настоящее время наиболее современной является версия 4.0, в которой помимо средств, унаследованных из предыдущих версий, заложены возможности придания Web-документам определенной динамики, например, динамического обновления части документа или управления пользователем элементами документа.

Спецификацией HTML предполагается, что размечаемый документ структурно делится на две части: «голову» (head) и «тело» (body). В «голове» документа указывается информация о документе, например, название, краткая аннотация, сведения об авторе и т. п. Содержимое этого раздела HTML-файла не отображается Web-клиентом вместе с основным текстом, а может быть доступно лишь частично и по требованию пользователя. В «теле» файла содержится основной текст документа вместе с разметкой, управляющей внешним представлением.

Принцип разметки с помощью HTML основан на использовании особых конструкций – *тегов*. В HTML *тег (tag, признак)* – специальное слово, заключенное в угловые скобки. Теги связываются с определенным фрагментом документа и указывают способ внешнего представления содержания этих фрагментов и их интерпретации Web-клиентом.

С функциональной точки зрения теги HTML можно распределить по следующим категориям:

- теги описания структуры документа и информации о нем, например, аннотации, списка ключевых слов и т. п.;
- теги, использующиеся для логической структуризации текста документа, например, для выделения заголовков, разбиения на абзацы, выделения цитат, создания списков, таблиц и т. п.;
- теги форматирования текста, описывающие параметры шрифтов, цвета и т. п.;
- теги организации гиперссылок;
- теги, устанавливающие связи HTML-файла с внешними объектами, например, графическими, звуковыми файлами и т. п.;

- теги создания форм, обеспечивающие возможность пользователю вводить информацию и передавать ее Web-серверу.
- Базовый вариант языка HTML обладает рядом недостатков:
- отсутствие средств, позволяющих динамически управлять внешним видом документа;
  - отсутствие средств, позволяющих легко менять внешнее оформление Web-документов без переделки самого документа, например, таблиц стилей;
  - отсутствие средств, позволяющих каким-либо образом структурировать содержание по смыслу, например, классифицировать понятия, встречающиеся в тексте.

В настоящее время существует ряд расширений HTML, а также набор различных технологий, позволяющих преодолеть недостатки базового варианта языка.

### 7.2.2. Таблицы стилей

*Таблицы стилей (каскадные таблицы стилей, Cascading Style Sheets, CSS)* – это отдельные файлы, которые содержат определение стилей оформления, применяемые к элементам содержания Web-документов. С помощью таблиц стилей можно определить, например, внешнее представление заголовков, основного текста документа, подписей к рисункам и т. п.

Создание таблицы стилей как отдельного файла позволяет использовать ее для описания внешнего вида сразу нескольких документов. При этом схема использования таблицы стилей выглядит следующим образом:

- создается файл таблицы стилей;
- в заголовке HTML-файла Web-документа с помощью специальных тегов указывается связь с файлом таблицы стилей;
- при разметке текста документа вместо прямого указания способа внешнего представления фрагментов указываются ссылки на элементы таблицы стилей;
- при необходимости изменения внешнего представления документа или его части, просто изменяются соответствующие элементы таблицы стилей.

### 7.2.3. Сценарии

*Сценарии (scripts)* являются еще одним расширением языка HTML. Сценарии представляют собой небольшие программы, текст которых включается в голову HTML-файла с помощью специальных тегов. При загрузке документа Web-клиент выделяет сценарии и передает их специальному модулю – интерпретатору языка, на котором написан сцена-

рий, для проверки правильности программного кода и его подготовки к последующему исполнению. Исполнение программы может осуществляться непосредственно при загрузке документа или же при выполнении определенных условий (например, при поступлении команды пользователя) – это определяется самим сценарием.

Сценарии широко используются для повышения функциональности документа, поскольку позволяют:

- придать динамику элементам Web-документа, например, организовать движение картинок;
- организовать динамическое обновление части Web-документа в зависимости от определенных условий, например, через определенные промежутки времени или по командам пользователя;
- организовать интерактивное взаимодействие с пользователем путем обработки определенных событий, например перемещения указателя «мыши» или нажатия каких-либо клавиш на клавиатуре.

#### **7.2.4. Язык *Dynamic HTML***

*Dynamic HTML* (*динамический HTML, DHTML*) – это расширение языка HTML, позволяющее создавать таблицы стилей и включать в состав Web-документа небольшие программы-сценарии (*script*), которые исполняются Web-клиентом и придания ему динамического характера, например движения картинок по документу. Для создания подобных сценариев используются специальные языки программирования. Наиболее распространенными являются два: язык *JavaScript*, разработанный компанией Netscape, и язык *VBScript* – разработка фирмы Microsoft. Для включения сценариев в документ используются специальные теги.

#### **7.2.5. *Java-апплеты***

*Java-апплет* – это небольшая программа, написанная на языке программирования Java. Подобные программы используются для придания динамического характера Web-документу. Классическим примером использования Java-апплетов является придание статическим картинкам определенных эффектов (падающего снега, движения волн по поверхности воды и т. п.), а также различные способы анимации динамически задаваемых текстовых надписей.

Язык программирования Java обладает рядом особенностей, упрощающих его использование в Интернет. К их числу следует отнести следующие два:

1. Исходный текст программы преобразуется не в машинные команды, а в специальный код, который не может напрямую исполняться процессором. Это обеспечивает аппаратную независимость про-

граммы и позволяет использовать ее на компьютерах различных типов. Однако такой подход требует наличия на компьютере, где исполняется Java-апплет, специального модуля (так называемой «Java-машины»), обеспечивающего преобразование независимого кода в машинные команды и их выполнение процессором.

2. В языке отсутствуют средства, позволяющие организовать прямое взаимодействие с устройствами и, прежде всего, с оперативной памятью. Это исключает возможность появления в тексте программ большого числа ошибок, которые могут привести к сбою компьютерной системы, а, следовательно, обеспечивает повышенную стабильность работы программ.

Java-апплеты реализуются в виде отдельных файлов и хранятся на Web-серверах. Возможность использования апплетов в Web-документах обеспечивается тем, что:

- на клиентском компьютере имеется программный компонент «Java-машина», обеспечивающий исполнение апплета (Java-машина может включаться либо в состав Web-клиента, либо – в состав операционной системы, управляющей работой Web-клиента);
- в состав языка HTML включен специальный тег, позволяющий подключить апплет к Web-документу и указать его адрес в Сети и входные параметры.

### 7.2.6. *Технология ActiveX*

*Технология ActiveX* – это еще одно средство расширения возможностей WWW. Принцип этой технологии основан на том, что для каждого типа информационных объектов (например, HTML-документов, документов Microsoft Office и т. п.) создаются специальные управляющие элементы.

*Управляющий элемент ActiveX (ActiveX-control)* – это программный модуль, обеспечивающий корректное отображение объекта и интерфейс, позволяющий пользователю работать с этим объектом. Управляющие элементы не могут функционировать как самостоятельные независимые программы. Они выполняются в пределах некоторой программной оболочки – ActiveX-контейнера. Важно отметить, что активизация и выполнение управляющего элемента происходит только при обращении к соответствующему объекту.

Такой подход позволяет строить программные средства, способные путем использования управляющих элементов ActiveX, осуществлять работу с большим количеством различных информационных объектов, а также с составными документами, объединяющими внутри себя различные объекты. В частности, именно благодаря этой технологии име-

ется возможность «встраивать» в Web-документы таблицы, подготовленные в Microsoft Excel или слайды презентаций Microsoft PowerPoint.

Возможность использования технологии ActiveX в WWW обеспечивается двумя механизмами:

- Web-клиент, поддерживающий технологию ActiveX, реализуется как контейнер, позволяющий при обращении к тому или иному информационному объекту динамически встраивать и выполнять «внутри себя» соответствующий управляющий элемент ActiveX;
- в состав языка HTML включен специальный тег, позволяющий устанавливать связи с объектами, доступными с помощью ActiveX.

Технология ActiveX – это общая технология построения программных комплексов. Она не является специализированной технологией World Wide Web. WWW – одна из областей применения ActiveX.

### **7.2.7. Язык XML**

*Расширяемый язык разметки XML (eXtensible Markup Language)* – это язык описания документов, во многом похожий на язык разметки гипертекста HTML, повсеместно используемый для конструирования Web-страниц. Однако он является гораздо более универсальным, чем HTML.

HTML – это язык, используемый для создания Web-страниц и основанный на определенном наборе «тегов», показывающих читающему текст программному обеспечению («браузеру»), как представлять содержимое страницы. Подобно HTML, XML представляет собой систему тегов, описывающих компоненты документа.

Язык XML позволяет описывать данные произвольного типа и используется для представления специализированной информации, например химических, математических, физических формул, медицинских рецептов, нотных записей, и т. д. Это означает, что XML может служить мощным дополнением к HTML для распространения в Web «нестандартной» информации.

Тело документа XML состоит из *элементов разметки (markup)* и непосредственно содержимого документа – *данных (content)*. XML – тэги предназначены для определения элементов документа, их атрибутов и других конструкций языка.

## **7.3. Унифицированные указатели ресурса**

### **7.3.1. Схемы и механизмы доступа к ресурсу**

Для обеспечения адресации Web-документов и других информационных объектов в свое время была предложена концепция унифицированных указателей ресурса *URL (Uniform Resource Locator)*.

URL представляет собой набор информации, необходимый для того, чтобы определить:

- узел сети, на котором расположен информационный объект;
- расположение информационного объекта на узле;
- метод получения доступа к объекту.

Унифицированный указатель ресурса (URL) представляет собой частный случай унифицированного определителя ресурса (Uniform Resource Identifier, URI). URI – является абстрактным понятием, описывающим принципы универсальной адресации ресурсов сети, позволяющие идентифицировать любой информационный объект в сети, независимо от его внутренней структуры (форматов представления данных, кодировок и т. п.), расположения и сетевых сервисов, обеспечивающих доступ к нему. В настоящее время определены различные подходы к реализации такой универсальной адресации. Одним из таких подходов является URL, для которого ключевой информацией для идентификации объекта является его местоположение (location) в сети.

Другой способ универсальной идентификации определен концепцией унифицированных имен ресурсов (Uniform Resource Name, URN), в которой ключевой информацией является уникальное мнемоническое имя ресурса, возможно никак не связанное с его физическим местоположением. Например, для однозначной идентификации человека в России можно использовать комбинацию «Фамилия-Имя-Отчество-Адрес Прописки» (аналог URL), или комбинацию «Номер и Серия паспорта» (аналог URN).

Поскольку для получения доступа к объекту все-таки необходимо знать его расположение (чтобы правильно адресовать запрос), то использование схем, не основанных на информации о расположении объекта, практически не очень удобно и требует дополнительных преобразований. Поэтому URL является наиболее популярной схемой адресации объектов в сети

Общая структура URL выглядит следующим образом:

### *СХЕМА: ЧАСТЬ-ЗАВИСЯЩАЯ-ОТ-СХЕМЫ*

Часть *СХЕМА* определяет способ получения доступа к ресурсу и то, каким образом представляется местоположение ресурса в части «*ЧАСТЬ-ЗАВИСЯЩАЯ-ОТ-СХЕМЫ*». Ниже приведены примеры URL:

**<http://www.site.ru>**

**<mailto:name@youraddress.ru>**

Как видно из примера, схема доступа может определять как конкретный протокол, так и целый сервис. В настоящее время существует целый ряд схем, которые могут использоваться в URL. Ниже перечислены наиболее часто используемые схемы:

Таблица 4

*Схемы и механизмы доступа к ресурсам*

Схема	Определяемый механизм доступа
mailto	Электронная почта
news	Телеконференции
ftp	Протокол FTP
telnet	Протокол Telnet
http	Протокол HTTP
https	Протокол HTTPS
file	Файл, расположенный на локальном компьютере

Независимо, от того, какая схема используется, «ЧАСТЬ-ЗАВИСЯЩАЯ-ОТ-СХЕМЫ» должна соответствовать следующим правилам:

- набор символов ограничен латинскими буквами, цифрами, символами «-» и «\_»;
- символы «%», «/», «:», «;», «.», «?», «@», «&», «#», «=», «+» используются как служебные (например, в качестве разделителей имен, входящих в состав адреса);
- не существует различия между прописными и строчными буквами.

### 7.3.2. Схема «mailto:»

Основная область применения схемы «mailto:» – создание сообщений электронной почты, адресованных определенным абонентам. Схемой определяется следующая структура URL:

1. Часть URL «АДРЕС» строится в соответствии с правилами построения адресов Электронной почты.
2. Часть «Параметры сообщения» позволяет заполнить основные поля заголовка почтового сообщения и указать текст, помещаемый в тело сообщения. При необходимости можно задать последовательно несколько параметров. Для объединения параметров используется символ «&».

Примеры URL:

- mailto:center@fio.ru – адресует одиночный почтовый ящик;
- mailto:center@fio.ru;fio@fio.ru – адресует одновременно два почтовых ящика;

- <mailto:center@fio.ru?subject=Question> – создает сообщение с заполненным полем Subject («Тема сообщения»);
- <mailto:center@fio.ru?cc=fio@fio.ru&body=Test> – создает сообщение с заполненным полем cc («Копия») и телом сообщения (параметр body).

### 7.3.3. Схема «ftp:»

Схема «ftp:» применяется для адресации файлов, доступ к которым может осуществляться по протоколу FTP. Для этой схемы предполагается следующая структура URL:

1. Имя Пользователя и Пароль – информация о пользователе, позволяющая проверить права доступа к ресурсу. Эта часть опускается, если осуществляется анонимный доступ.
2. Адрес Узла – полное доменное имя (FQDN) или IP адрес узла, на котором расположен ресурс.
3. Порт – номер TCP-порта. Этот параметр может быть опущен, если FTP-сервер, работающий на указанном узле, использует стандартный порт.
4. Путь – путь (последовательность каталогов) внутри узла, определяющий расположение объекта. В качестве разделителя используется символ «/».
5. Имя Файла – имя файла, к которому осуществляется доступ. Если имя файла не опущено, то такой URL адресует каталог, указанный в части «Путь». По запросу с таким адресом клиенту будет передано оглавление указанного каталога.
6. Тип Ресурса – эта часть указывает, является ли адресуемый файл текстовым или нет, что необходимо для выбора режима передачи файла. Этот параметр является необязательным и может опускаться.

Поскольку для сервиса FTP информационным объектом является статический файл независимо от его формата и содержания, то обращение по схеме «ftp:» не предполагает механизм адресации фрагмента файла, а также передачу параметров для генерации объектов.

Примеры URL для схемы «ftp:»:

- <ftp://center.fio.ru/> или <ftp://213.128.193.154/> – адресует корневой каталог FTP-сервера;
- <ftp://center.fio.ru:1/progr/readme.txt> – адресует определенный файл; при этом указывается, что FTP-сервер использует нестандартный порт;
- <ftp://guest:qwerty@213.128.193.154/readme.txt> – адресует определенный файл, для неанонимного доступа; доступ осуществляется от имени пользователя «guest» с паролем «qwerty».

### 7.3.4. Схемы «http:» и «https:»

Схемы «http:» и «https:» обычно используется для адресации к Web-документам и некоторым другим объектам, например, графическим и звуковым файлам, доступным по протоколам HTTP и HTTPS. Эти схемы предполагают следующую структуру URL:

1. Имя Пользователя и Пароль – информация о пользователе, позволяющая проверить права доступа к ресурсу. Эта часть опускается, если осуществляется анонимный доступ.
2. Адрес Узла – полное доменное имя (FQDN) или IP адрес узла, на котором расположен ресурс;
3. Порт – номер TCP-порта. Этот параметр может быть опущен, если HTTP-сервер, работающий на указанном узле, использует стандартный порт.
4. Путь к Ресурсу – путь (последовательность каталогов) внутри узла, определяющий расположение объекта. В качестве разделителя используется символ «/».
5. Файл Ресурса – имя файла, содержащего информационный объект, например, Web-документ. В некоторых случаях имя файла и путь к ресурсу могут отсутствовать. В этом случае Web-сервер будет выдавать документ, имя которого задано у него в настройках как имя «по умолчанию». Наиболее часто в качестве имени по умолчанию используются следующие варианты: «index.htm», «index.html», «default.htm» или «default.html». Конкретный вариант настройки определяется администратором Web-сервера.
6. Параметры Запроса – список параметров и их значений, которые предназначены для динамической обработки запроса. Если адресуемый документ не содержит внутри себя программного кода, а также не формируется сервером динамически, то эта часть опускается. При необходимости можно задать последовательно несколько параметров, используя для объединения символ «&».
7. Часть Ресурса – содержит имя закладки («якоря», anchor), связанной с некоторым фрагментом внутри Web-документа. Использование этого параметра позволяет адресовать часть Web-документа. Следует отметить, что адресация части документа позволяет отобразить документ в окне браузера, начиная с позиции, с которой связана закладка, а вовсе не означает, что файл будет передаваться клиенту частично.

Как уже отмечалось, некоторые компоненты этой конструкции могут опускаться. В большинстве случаев для описания URL достаточно упрощенных вариантов.

Примеры URL для схемы «http:»:

- <http://www.fio.ru> или <http://213.128.193.154> – адресует страницу по умолчанию (index.html);
- <http://center.fio.ru/workroom/Ruslang/list2.htm> – адресует определенную страницу;
- [http://213.128.193.154/image/blank\\_ls.gif](http://213.128.193.154/image/blank_ls.gif) – адресует графический файл;
- <http://center.fio.ru/default.asp#top> – адресует часть документа;
- [http://center.fio.ru/FORUM\\_MCIO/ReadTopics.asp?Topicid=32&FORUMID=23](http://center.fio.ru/FORUM_MCIO/ReadTopics.asp?Topicid=32&FORUMID=23) – адресует страницу, динамически создаваемую сервером по запросу.

## **7.4. Клиентское программное обеспечение**

Одной из главных концепций WWW является концепция «универсальной читаемости» (Universal readership), т. е. обеспечение возможности получения доступа и организации работы с информационными объектами различных типов с помощью одной клиентской программы. Именно поэтому с самого начала на клиентское программное обеспечение накладывалось требование универсальности, как с точки зрения представления пользователю различных информационных объектов (Web-документов, тестовых файлов, графических файлов и т. п.), так и с точки зрения обеспечения доступа к различным серверам по различным протоколам (например, к FTP-серверам).

Естественно, что с момента появления сам по себе сервис WWW постоянно развивается, и, следовательно, меняется состав функций, которые должны выполнять Web-клиенты.

Функции современных Web-клиентов можно разделить на две группы: стандартные функции, которые поддерживаются всеми клиентами и обеспечивают основные механизмы работы с WWW, и дополнительные, которые в принципе являются не обязательными и поддерживаются всеми развитыми клиентами для повышения удобства работы пользователей.

## **7.5. Стандартные функции Web-клиентов**

### ***7.5.1. Доступ к различным информационным ресурсам***

Изначально концепцией WWW было предусмотрено, что Web-клиент является универсальным клиентом, обеспечивающий доступ к различным программным серверам по протоколам HTTP, FTP и NNTP.

В настоящее время Web-клиенты позволяют непосредственно осуществлять доступ к Web-серверам (протоколы HTTP и HTTPS), FTP-серверам (протокол FTP). При необходимости обращения к ресурсам, доступным через другие протоколы, Web-клиенты передают управление соответствующим

щим программным клиентам, которые могут быть реализованы либо как отдельные приложения, либо как дополнительно встраиваемые программные расширения Web-клиентов (plugins). Такая программная интеграция позволяет осуществлять доступ к ресурсам следующих сервисов:

- WWW;
- FTP;
- Электронная почта;
- Телеконференции;
- Telnet;
- мультимедиа ресурсы (например, RealAudio и RealVideo).

### ***7.5.2. Работа с объектами различного типа***

Универсальность Web-клиента предполагает возможность работы с различными информационными объектами. Сам по себе термин «работа с объектами» в общем случае предполагает выполнение следующих действий:

- получение по запросу пользователя объекта от сервера;
- корректное отображение объекта, печать и обеспечение интерактивного взаимодействия с объектом, если это требуется;
- сохранение объекта в виде отдельного файла или набора файлов на диске;
- перенос объекта из окна клиента в другое приложение через буфер обмена.

Наиболее сложным для практической реализации является отображение объекта и обеспечение интерактивного взаимодействия с ним. Естественно, что на практике невозможно создать программный продукт, который позволит работать с объектами любого типа. Поэтому при разработке Web-клиентов обычно реализуется трехуровневый подход обеспечения универсальности работы:

1. Первый уровень предполагает, что ряд информационных объектов отображается и управляется непосредственно Web-клиентом. К числу таких объектов относятся:
  - HTML-файлы;
  - графические изображения в форматах GIF, JPEG и PNG, входящие в состав Web-документа (или как независимые файлы);
  - неформатированный текст.
2. Второй уровень предполагает, что часть объектов с помощью дополнительных программных расширений Web-клиента (plugins). Примером таких объектов являются объекты Macromedia Flash.
3. Третий уровень предполагает, что для ряда объектов Web-клиент не может обеспечить полноценную обработку. В этом случае мо-

жет быть привлечено дополнительное программное обеспечение, специально предназначенное для работы с такими объектами, либо объекты могут быть просто сохранены в виде отдельного файла или набора файлов на диске пользовательского компьютера. В качестве примера можно привести документы Microsoft Office.

### ***7.5.3. Обеспечение навигации***

Механизм навигации предполагает обеспечение возможности перехода от одного информационного объекта к другому, т. е. своеобразного «путешествия» по информационному пространству. Современные клиенты позволяют выполнять такое перемещение несколькими способами:

- путем непосредственного задания адреса объекта;
- путем перехода по гиперссылкам от одного к другому;
- путем отката назад по «пройденному пути» и перемещения вперед после отката;
- путем использования «истории» или «журнала», где регистрируются последние посещения.

### ***7.5.4. Кэширование и управление загрузкой объектов***

Работа с любым объектом предполагает, прежде всего, его передачу по сети от сервера к клиенту, а это требует определенного времени. Поэтому современные клиенты обеспечивают кэширование объектов, то есть временное сохранение полученных объектов на локальном диске пользовательского компьютера. Это ускоряет их повторное использование, поскольку не требуется при вторичном обращении получать объект с сервера. Единственное, что может потребоваться – это проверка факта изменения объекта, которое могло произойти после его кэширования.

Другим механизмом ускорения работы является управление загрузкой объектов. Это, прежде всего, относится к Web-документам. Поскольку в общем случае Web-документ представлен целым комплексом файлов: основным HTML-файлом, файлами иллюстраций и т. д., то его загрузка состоит из нескольких запросов – по одному на каждый составляющий файл. Для ускорения работы Web-клиенты обеспечивают режимы, в которых пользователь может отказаться от автоматической загрузки каких-либо составляющих, например, графических файлов.

## **7.6. Дополнительные функции Web-клиентов**

### ***7.6.1. Поддержка маркеров «cookie»***

Протокол HTTP, который в основном используется для взаимодействия между Web-клиентом и Web-сервером, не позволяет установ-

ливать длительного сеанса. Это означает, что при поступлении запроса от определенного клиента сервер не имеет возможности «понять», поступали ли ранее запросы от данного клиента или нет. Другими словами взаимодействие не предусматривает механизма идентификации клиента сервером.

В ряде случаев это может оказаться неудобным. Например, при построении Интернет-магазина серверу необходимо при переходе клиента с одной страницы на другую «помнить» информацию о клиенте и об отобранных им товарах.

Для решения этой проблемы в свое время был предложен механизм, основанный на использовании специальных *маркеров cookie*.

*Маркер cookie* – это небольшой файл, который сохраняется Web-клиентом на локальном диске по запросу Web-сервера, содержащий информацию о просмотренных страницах, о клиенте, а также настройках и параметрах, установленных пользователем для этих страниц. При повторном обращении информация, сохраненная в соответствующем файле cookie, передается серверу.

Современные Web-клиенты поддерживают работу с маркерами cookie в различных режимах. Возможны следующие варианты:

- Web-клиент автоматически сохраняет файлы cookie по запросу сервера;
- Web-клиент выборочно сохраняет файлы cookie по какому-либо условию, например, по разрешению пользователя или только связанные с определенными узлами;
- Web-клиент не сохраняет файлы cookie ни при каких обстоятельствах.

### ***7.6.2. Поддержка Java-апплетов и технологии ActiveX***

Для обеспечения работы с Java-апплетами и объектами ActiveX Web-клиент должен отвечать определенным требованиям:

- в составе Web-клиента должна входить «Java-машина», обеспечивающая исполнение Java-апплетов, либо Web-клиент должен уметь взаимодействовать с внешней Java-машиной, например, входящей в состав операционной системы;
- для поддержки технологии ActiveX Web-клиент должен быть реализован как ActiveX-контейнер, обеспечивая тем самым функционирования управляющих элементов ActiveX.

Поскольку поддержка Java-апплетов и технологии ActiveX связана с исполнением внешних программ, полученных из сети Интернет, то Web-клиент должен с целью обеспечения безопасности пользовательского узла контролировать действия и ограничивать возможности этих

программ. Современные Web-клиенты позволяют работать с Java-апплетами и управляющими элементами ActvaX в следующих режимах:

- в режиме, когда функционирование программного объекта позволяет;
- в режиме, когда работа программного объекта в целом или выполнение определенных действий блокируется;
- в режиме, когда Web-клиент запрашивает у пользователя возможность выполнения определенных действий или функционирования объекта в целом.

## **7.7. Протоколы взаимодействия клиента и сервера**

Основным протоколом взаимодействия WWW-клиента и WWW-сервера является протокол HTTP (Hypertext Transfer Protocol). Этот протокол обычно использует в качестве средства доставки своих сообщений транспортный протокол TCP.

Следует отметить, что спецификацией HTTP предусмотрено использование любого транспортного протокола, обеспечивающего гарантированную доставку

Протокол предполагает, что передаваемая информация представляется исключительно в виде текста в формате ASCII. Это оказывается не удобным для передачи данных различного вида – текста, написанного с помощью нелатинских букв, графических, звуковых, виде файлов и т. п. Для решения этой проблемы используются методы кодирования MIME, позволяющие преобразовать данные любого вида к тексту ASCII.

Принцип взаимодействия WWW-клиента и WWW-сервера на основе протокола HTTP предполагает следующую последовательность действий:

- клиент формирует сообщение-запрос и передает серверу;
- сервер получает сообщение, анализирует и обрабатывает запрос, формирует сообщение-ответ и направляет его клиенту.

В сообщениях-запросах в текстовом виде указывается строка запроса, содержащая команду, которую должен выполнить HTTP-сервер, а также параметры запроса, например URL, определяющий объект, над которым выполняются действия. Команды протокола позволяют клиенту:

- запросить у сервера ресурс (например, Web-документ, графический файл и тип) или проверить возможность доступа к ресурсу;
- передать данные серверу (например, содержимое заполненных полей формы);
- удалить ресурс на сервере;
- установить или отменить связь ресурса с другими ресурсами.

Как правило, для того, чтобы получить и представить пользователю запрашиваемый Web-документ, клиенту требуется выполнить несколько запросов. Это объясняется тем, что Web-документ, в общем

случае представляется целым набором файлов, а для получения каждого файла требуется выполнить отдельный запрос.

Сообщения-ответы содержат в себе строку состояния, определяющую результат выполнения сервером команды клиента. Эта строка состоит из двух частей – числового кода и текстового описания результата. Например, строка сообщает клиенту о том, что запрашиваемый ресурс не найден.

Поскольку протокол предполагает передачу данных в текстовом виде (ASCII), то заголовок представляет последовательность строк.

В полях заголовка передается информация, необходимая для корректного получения и правильной интерпретации Web-документов и других ресурсов, которые передаются с помощью протокола:

- данные о клиенте, например, название, версия, поддерживаемые кодировки и т. п. и о сервере;
- данные о пользователе (при необходимости), который запрашивает ресурсы, например, информация для аутентификации и адрес электронной почты;
- информация о передаваемом ресурсе, например, тип содержимого (текст, графика и т. п.), язык, размер данных, дата обновления и т. п.;
- информация о способе кодирования данных при передаче, например, версия MIME и использование алгоритмов сжатия данных.

Состав заголовка HTTP-сообщения зависит от того, является ли сообщение запросом клиента или ответом сервера, а также от типа запроса или ответа.

### **Методические указания**

В связи с ограничением объема пособия, в данном разделе изложены только общие принципы организации сервиса World Wide Web (WWW):

- сервис WWW предназначен для создания интегрированного информационного пространства, предусматривающего установление взаимосвязей между различными информационными объектами (текстовыми, графическими и звуковыми файлами, сообщениями и т. п.);
- основой сервиса WWW являются гипертекст и гиперссылки, когда в тексте статьи предусмотрен механизм «отсылки» (например, выделением терминов в тексте статьи) читателя к другим статьям, связанным с ней по смыслу;
- сервис WWW включает в себя следующие компоненты: Web-документ, Web-сайт, Web-сервер, Web-клиент и протокол HTTP;
- Web-документ (Web-страница) – гипертекстовый документ, содержащий в себе гиперссылки на другие Web-документы, различные информационные объекты;

- Web-сайт – совокупность объединенных по смыслу и связанных с помощью гиперссылок Web-документов;
- Web-сервер – серверное программное обеспечение, предназначенное для управления и пересылки по запросу Web-документов клиентам;
- Web-клиент (браузер) – клиентская программа, позволяющая получить Web-документ или другой объект, представить его пользователю и обеспечивающая возможность работы с ним;
- протокол HTTP (Hyper Text Transfer Protocol) – протокол взаимодействия клиента с Web-сервером;
- для описания внешнего вида Web-документа и его связей с другими документами и объектами разработан специальный язык разметки гипертекста HTML (Hyper Text Markup Language);
- в основе языка HTML лежит принцип использования особых конструкций – «тегов», представляющих собой заключенные в угловые скобки специальные слова, которые связаны с определенными фрагментами документа и указывают способ внешнего представления содержания этих фрагментов и их интерпретации Web-клиентом;
- таблицы стилей (каскадные таблицы стилей, Cascading Style Sheets, CSS) – отдельные файлы, с помощью которых можно определить внешнее представление фрагментов Web-документа (например, заголовков, основного текста документа, подписей к рисункам и т. п.);
- сценарии (scripts) – небольшие программы, которые используются для повышения функциональности Web-документа (например, движения картинок, обновление части Web-документа через определенные промежутки времени или по командам пользователя и т. п.);
- Dynamic HTML (динамический HTML, DHTML) – расширение языка HTML, позволяющее определять таблицы стилей и включать в состав Web-документа небольшие программы-сценарии (script), созданные с помощью специальных языков программирования JavaScript и VBScript;
- Java-апплет – программа, написанная на языке программирования Java и обеспечивающая возможность придания динамического характера статическим картинкам Web-документа (например, эффектов падающего снега, движения волн по поверхности воды и т. п.);
- технология ActiveX позволяет «встраивать» в Web-документы различные информационные объекты (например, таблицы, подготовленные в Microsoft Excel или слайды презентаций Microsoft PowerPoint);
- управляющий элемент технологии ActiveX (ActiveX-control) – программный модуль, обеспечивающий корректное отображение информационного объекта и интерфейс пользователю для работы с этим объектом;

- расширяемый язык разметки XML (eXtensible Markup Language) позволяет описывать данные произвольного типа и используется для представления специализированной информации (например, химических, математических, физических формул, медицинских рецептов, нотных записей, и т. д.);
- унифицированный указатель ресурса URL (Uniform Resource Locator) позволяет идентифицировать узел сети, на котором расположен информационный объект, и метод получения доступа к объекту;
- схема URL определяет механизм доступа к ресурсам: telnet – протокол Telnet, ftp – протокол FTP, news – телеконференции, mailto – электронная почта, http – протокол HTTP;
- программное обеспечение Web-клиента обеспечивает доступ к ресурсам следующих сервисов: Telnet, FTP, Usenet, E-mail, WWW.

## ГЛАВА 8. СЕРВИС SE – ПОИСКОВЫЕ СИСТЕМЫ

### 8.1. Назначение сервиса SE

*Сервис SE (Search Engine) – поисковые системы* – это специальные Web-сайты, на которых пользователь по заданному запросу может получить ссылки на сайты, соответствующие этому запросу. Установлено, что 85 % пользователей Интернет используют поисковые системы, чтобы найти необходимые им товары, услуги и информацию.

Поисковые системы отличаются от тематических каталогов. Они представляют собой серверы с огромной базой URL-адресов, которые автоматически обращаются к Web-страницам по этим адресам, изучают содержимое этих страниц, формируют и прописывают ключевые слова со страниц (индексируют страницы). Более того, эти серверы обращаются по всем встречаемым на страницах ссылкам и, переходя к новым страницам, проделывают с ними то же самое. Так как почти любая Web-страница имеет множество ссылок на другие страницы, то при подобной работе поисковая система в конечном результате теоретически может обойти все сайты в Internet.

Поисковая система состоит из следующих основных компонентов:

1. *Паук (spider)* – программа, которая скачивает Web-страницы тем же способом, что и браузер пользователя. Отличие состоит в том, что браузер отображает информацию, содержащуюся на странице (текстовую, графическую и т. д.), паук же не имеет никаких визуальных компонент и работает напрямую с html-текстом страницы (можно сделать «просмотр html-кода» в браузере, чтобы увидеть «сырой» html-текст).
2. *Путешествующий паук (crawler)* – программа, которая выделяет все ссылки, присутствующие на странице. Ее задача – определить, куда дальше должен идти паук, основываясь на ссылках или исходя из заранее заданного списка адресов. Краулер, следуя по найденным ссылкам, осуществляет поиск новых документов, еще неизвестных поисковой системе.
3. *Индексатор (indexer)* – программа, которая разбирает страницу на составные части и анализирует их. Выделяются и анализируются различные элементы страницы, такие как текст, заголовки, структурные и стилевые особенности, специальные служебные html-теги и т. д. Результатом анализа является index-файл.
4. *База данных (database)* – это хранилище всех index-файлов, полученных поисковой системой в процессе скачивания и анализа

Web-страницы. Иногда базу данных называют индексом поисковой системы.

5. *Система выдачи результатов (search engine results engine)* – занимается ранжированием страниц. Она решает, какие страницы удовлетворяют запросу пользователя, и в каком порядке они должны быть отсортированы. Это происходит согласно алгоритмам ранжирования поисковой системы. С этим компонентом поисковой системы взаимодействует оптимизатор, пытаясь улучшить позиции сайта в выдаче с помощью определенных факторов, влияющих на ранжирование результатов.
6. *Web-сервер (Web – server)* – сервер, который осуществляет взаимодействие между пользователем и остальными компонентами поисковой системы. Как правило, на сервере присутствует html-страница с полем ввода, в котором пользователь может задать интересующий его поисковый термин. Web-сервер также отвечает за выдачу результатов пользователю в виде html-страницы.

Детальная реализация поисковых механизмов может отличаться друг от друга. Например, связка *spider+crawler+indexer* может быть выполнена в виде единой программы, которую называют *поисковым роботом*. Она скачивает известные Web-страницы, анализирует их, ищет по ссылкам новые ресурсы, индексирует их и заносит в базу данных в виде *index-файла*. Поиск осуществляется другой программой, которая извлекает запрашиваемую пользователем информацию из *index-файла*. Однако всем поисковым системам присущи описанные общие черты.

## 8.2. Критерии ранжирования документов

В поисковых системах ссылки на документы сортируются (ранжируются) по мере соответствия запросу. Для ранжирования страниц в поисковой выдаче используются *текстовые критерии, ссылочные критерии и критерии пользовательской оценки*.

*Текстовые критерии* определяют релевантность документа по совпадению слов и их сочетаний в запросе и тексте и заголовке страницы.

*Релевантность документа* – показатель, отражающий соответствие содержания документа конкретному запросу поисковой системы. Поисковые системы рассчитывают релевантность документа, строя частотный ряд из встречающихся на странице слов и словосочетаний. Чем чаще они встречаются в документе, тем большую по отношению к запросу пользователя релевантность он получает.

Поисковые системы отображают ссылки на Web-страницы документов в порядке убывания релевантности частями по 10 – 20 ссылок. Согласно данным маркетинговых исследований около 60 % пользователей

ограничиваются первой страницей результатов поиска и почти 90 % – первыми тремя страницами. Отсюда следует задача для специалистов по *сайтпромоутингу* – добиться, чтобы независимо от построения запроса страницы Web-сайта стояли в первых 10-20 результатах поиска.

Основные текстовые критерии ранжирования документов приведены в табл. 5.

Таблица 5

*Текстовые критерии ранжирования документов*

Критерии	Логика ранжирования
«Вес» слова	Чем выше частота повторения слова в документе, тем больше ранг документа
Взаимное положение слов	Учет полного совпадения фраз или их подобия (например, порядок и близость слов друг к другу)
Положение найденного текста по отношению к началу документа	Считается, что чем ближе расположена информация к началу документа, тем выше ее значение
Наличие слов запроса в выделенных фрагментах и заголовках	Значимость обнаружения искомого текста в выделенных фрагментах считается выше, чем в обычном тексте
Совпадение темы страницы с темой запроса	Использование в поиске слов, не содержащихся в тексте запроса, но соответствующих теме запроса
Совпадение названия домена или файла с ключевым словом	Поисковые машины придают дополнительный «вес» страницам, у которых домен или имя файла совпадают с ключевым словом
Совпадение поискового запроса с описанием из каталога	Сайт получает более высокий рейтинг, если слова поискового запроса совпадают с описанием каталога поисковой системы
Значимость редких слов	Значимость каждого из поисковых слов тем больше, чем реже оно встречается в документе

Оценку значимости фрагментов текста выработал Г. Лун. Он предложил оценивать фрагменты текста по следующему выражению:

$$V = \frac{N_k^2}{N_o}$$

где:  $V$  – значимость фрагмента;  $N_k$  – число ключевых слов в данном фрагменте;  $N_o$  – общее число слов во фрагменте.

Система выявления ключевых слов обычно использует статистический частотный анализ (методика В. Пурто). Пусть:

$F$  – частота, с которой встречаются различные слова в тексте;

$P$  – относительное значение полезности (важности);

$C$  – константа, которая определяет соотношение частоты слов и их полезности.

Тогда зависимость  $F(P)$  определяется формулой:

$$F(P) = C \frac{1}{P}.$$

Данное положение предполагает существование двух граничных значений частот:

- слова с частотой менее нижней границы считаются слишком редкими (не способными отразить смысл документа), а с частотой, превосходящей верхнюю границу, считаются общими, не несущими смысловой нагрузки;
- слова с частотой, находящейся между этими границами, в наибольшей степени характеризуют содержимое данного конкретного документа.

Согласно *ссылочным критериям* документ ранжируется с учетом индекса цитирования.

*Индекс цитирования* – это показатель известности сайта в Интернете, определяемый числом и значимостью ссылок на других сайтах на искомый ресурс. Общее число внешних ссылок на сайт не подходит в качестве критерия для расчета цитируемости, т. к. значимость ссылок на непопулярных ресурсах ничтожна по сравнению со значимостью ссылок с известных сайтов.

При определении индекса цитирования учитывается не только число внешних ссылок на сайт, но индекс цитирования самих сайтов, ссылающихся на данный. Наиболее ценные ссылки – ссылки, размещенные на главной странице высокоцитируемых сайтов. В общем случае каждая прямая ссылка на Web-страницу увеличивает ее цитируемость на величину, пропорциональную цитируемости ссылающейся страницы и обратно пропорциональную общему числу ссылок на ссылающейся странице.

В свое время двумя американскими аспирантами Сергеем Брином и Ларри Пейджем, основавшими в 1997 году поисковую машину *Google*, была разработана модель, эмулирующая движение пользователя по документам в сети. При этом предполагалось, что пользователь с равной долей вероятности перейдет по любой из ссылок, содержащихся в документе, который он в данный момент просматривает. *Следова-*

тельно, вероятность пользователя попасть на конкретный документ зависит от количества ссылок на него с других документов и от того, насколько вероятно нахождение пользователя на одном из ссылающихся документов и сколько исходящих ссылок содержит этот документ. Эта вероятность и была принята за показатель авторитетности или ранг страницы (*PageRank*):

$$PR_a = (1 - d) + d \sum_{i=1}^n \frac{PR_i}{C_i},$$

где:  $PR_a$  – PageRank страницы  $a$ ;  $d$  – коэффициент затухания (означает вероятность того, что пользователь, зашедший на страницу, перейдет по одной из ссылок, содержащейся на этой странице, а не прекратит путешествие по сети), обычно устанавливают равным 0,85;  $i$  – страница, содержащая ссылки на страницу  $a$  ( $i$  изменяется от 1 до  $n$ );  $PR_i$  – PageRank страницы  $i$ , ссылающейся на страницу  $a$ ;  $C_i$  – общее число ссылок на странице  $i$ ;  $1 / C_i$  – вероятность того, что пользователь, находящийся на странице  $i$ , из  $C_i$  доступных ему ссылок выберет именно ссылку на страницу  $a$ ;  $d * PR_i / C_i$  – поток «теоретической посещаемости», который дойдет до страницы  $a$  со страницы  $i$  (суммирование идет по всем страницам, ссылающимся на страницу  $a$ );  $(1 - d)$  – минимальный PageRank страницы (он не равен нулю за счет того, что пользователь регулярно выбирает новый сайт в качестве стартовой точки).

Одним из распространенных заблуждений является то, что можно вычислить PageRank по этой формуле для отдельно взятого документа, используя известные значения PageRank для ссылающихся на него документов. Так делать нельзя. Чтобы вычислить PageRank какого-либо документа надо составить систему  $N$  линейных уравнений данного вида для каждого из документа из поисковой базы, где  $N$  – количество документов в поисковой базе. Причем, для выполнения условия, что сумма значений PageRank для всех документов (т. е. вероятность того, что пользователь находится на любой из страниц) равна 1, к свободный члену  $(1 - d)$  в каждом уравнении добавляют множитель  $1/N$ . Эта система будет содержать  $N$  неизвестных. Решив ее, получим значения PageRank для каждого документа, известного поисковой машине.

В поисковой базе крупнейших поисковых машин содержится огромное количество документов. Несмотря на то, что матрица, соответствующая системе уравнений будет сильно разрежена, численное

решение этой системы требует огромных вычислительных мощностей. Поэтому поисковая система должна постараться максимально упростить процесс расчета, вводя некоторые допущения. Вот эти конкретные особенности реализации классической формулы PageRank, увы, составляют коммерческую тайну поисковых машин.

Согласно критерию *пользовательской оценки* для ранжирования страниц в поисковой выдаче используются системы оценки качества страниц пользователями, которые основаны на предположении: если пользователь переходит по ссылке, значит он счел ее интересной, и если долго не возвращается на страницу поисковой системы, значит его ожидания подтвердились.

Поисковая система *Rambler* при ранжировании результатов поиска в ответах на поисковый запрос использует *коэффициент популярности*, определяемый числом пользователей, которые просматривали данную страницу за последние несколько недель. Данный коэффициент, как и алгоритм PageRank, основан на учете гиперссылок между страницами сети, однако эта реализация дополнительно использует данные о реальной посещаемости страниц, полученные от счетчика Top100. Дело в том, что «классические» ссылочные алгоритмы фактически учитывают мнение только одной категории пользователей сети – Web-мастеров. Действительно, если большому количеству Web-мастеров нравится тот или иной ресурс, они размещают на него ссылки. Обычные пользователи, как правило, созданием страниц и сайтов не занимаются, и поэтому учесть их мнение оказывается невозможно. Счетчик Top100 как раз и предназначен для того, чтобы сделать коэффициент популярности более справедливым.

Однако, судя по всему, в последнее время данные о посещаемости документов, полученные от счетчика Top100, оказывают все меньшее и меньшее влияние на коэффициент популярности, так как счетчик не в состоянии противостоять массовым накруткам, практикуемым владельцами некоторых сайтов. Соответственно, все большее значение приобретает составляющая, вычисляемая на основе учета гиперссылок между страницами сети.

В соответствии с изложенными критериями формулу, приближенно описывающую процесс определения релевантности документа запросу, можно представить следующим образом:

$$R_a(x) = (m * T_a(x) + p * L_a(x)) * F(PR_a),$$

где:  $R_a(x)$  – итоговое соответствие документа  $a$  запросу  $x$ ;  $T_a(x)$  – релевантность текста (кода) документа  $a$  запросу  $x$ ;  $L_a(x)$  – релевант-

ность текста с учетом ссылок с других документов на документ  $a$  запросу  $x$ ;  $PR_a$  – PageRank страницы  $a$ ;  $F(PR_a)$  – монотонно неубывающая функция, причем  $F(0)=1$  и можно допустить, что  $F(PR_a)=(1+q*PR_a)$ ;  $m, p, q$  – весовые коэффициенты, определяемые разработчиком поисковой системы.

Резюмируя вышеизложенное, можно отметить, что для повышения ранга страницы необходимо работать над тем, чтобы как можно большее количество документов сети ссылалось на нее. Делать это можно различными способами – с помощью обмена ссылками с другими сайтами, регистраций в каталогах и различных тематических ресурсах и т. д. Идеальный способ – сделать свой сайт настолько уникальным и интересным, чтобы владельцы других ресурсов сами считали необходимым поставить ссылку на него. Не следует также забывать, что при расчете ранга документа учитываются как внешние, так и внутренние ссылки. Поэтому грамотная перелинковка документов внутри сайта позволяет повысить ранг самых важных из них с точки зрения содержащейся информации. Наиболее важные в этом смысле документы обязательно должны иметь ссылку с главной страницы сайта, которая, как правило, имеет максимальный ранг среди всех страниц сайта вследствие того, что на нее указывает большинство внешних ссылок на сайт.

### **8.3. Организация поиска документов**

#### **8.3.1. Основные поисковые системы**

В настоящее время существует 3 основных международных поисковых системы: *Google, Yahoo и MSN Search*, имеющих собственные базы и алгоритмы поиска. Большинство остальных поисковых систем использует в том или ином виде результаты 3-х перечисленных. Например, поиск AOL ([search.aol.com](http://search.aol.com)) и Mail.ru используют базу Google, а AltaVista, Lycos и AllTheWeb – базу Yahoo.

В зоне русскоязычного Internet действуют более трех десятков поисковых систем. Причем около 90 % аудитории используют 3 самые популярные поисковые системы: Yandex, Google, Rambler (табл. 6).

Для поиска необходимых документов следует обратиться к конкретной поисковой системе и составить поисковый запрос, который может включать в себя одно или несколько слов. В запросе могут присутствовать знаки препинания. Составлять простые запросы можно и не вдаваясь в тонкости языка запросов. Так, если ввести в поисковую строку несколько слов без знаков препинания и логических операторов, то будут найдены документы, содержащие все эти слова (причем на ограниченном расстоянии

друг от друга). Однако знание и правильное применение языка запросов поисковой системы поможет сделать поиск быстрым и эффективным.

Таблица 6

*Поисковые системы русскоязычного Internet*

Поисковая система	Сервер	Доля аудитории
Яндекс	<a href="http://www.yandex.ru">http://www.yandex.ru</a>	до 67 %
Google	<a href="http://www.google.com">http://www.google.com</a>	до 33 %
Rambler	<a href="http://search.rambler.ru">http://search.rambler.ru</a>	до 20 %
Поиск@Mail.ru	<a href="http://go.mail.ru">http://go.mail.ru</a>	до 15 %
АПОРТ	<a href="http://www.aport.ru">http://www.aport.ru</a>	до 10 %
MSN	<a href="http://search.msn.com">http://search.msn.com</a>	до 3 %
Nigma	<a href="http://nigma.ru/">http://nigma.ru/</a>	до 3 %
Yahoo!	<a href="http://search.yahoo.com">http://search.yahoo.com</a>	менее 1 %
Altavista	<a href="http://www.altavista.com">http://www.altavista.com</a>	менее 1 %
WebAlta	<a href="http://webalta.ru/">http://webalta.ru/</a>	менее 1 %

Все поисковые системы используют сходные принципы языка запросов. Ссылку на полное описание языка запросов для каждой поисковой системы можно найти на ее главной странице. В большинстве языков запросов кроме простого запроса можно задавать операторы И (AND), ИЛИ (OR), НЕ (NOT), метасимвол \*, заменяющий до 5 произвольных символов, коэффициентные символы + и –, служащие для увеличения или уменьшения значимости вводимых в запросе слов.

### **8.3.2. Поисковая система Яндекс**

Доступ к поисковой системе *Яндекс* (<http://www.yandex.ru>) был открыт в 1997 году. Поиск осуществляется не только по Web-страницам, но и по специализированным массивам данных, среди которых новости ведущих информационных агентств, товары Internet-магазинов, ресурсы WAP –серверов.

Яндекс поддерживает собственный каталог Internet-ресурсов, формирующийся на основе *индекса цитирования* Яндex'a (CY – Citation Yandex). CY какой-либо Web-страницы измеряется количеством других страниц, содержащих ссылки на эту страницу. Этот метод оценки ресурсов принципиально отличается от простого учета количества посещений страницы.

Яндекс имеет *простой и расширенный поисковые интерфейсы*, а также страницу настройки формата выдачи результатов поиска. Он использует собственную систему обозначений логических операторов, а также поддерживает большое количество поисковых функций.

В верхней части домашней страницы поисковой системы Яндекс расположено поле для ввода ключевых слов (рис. 11).

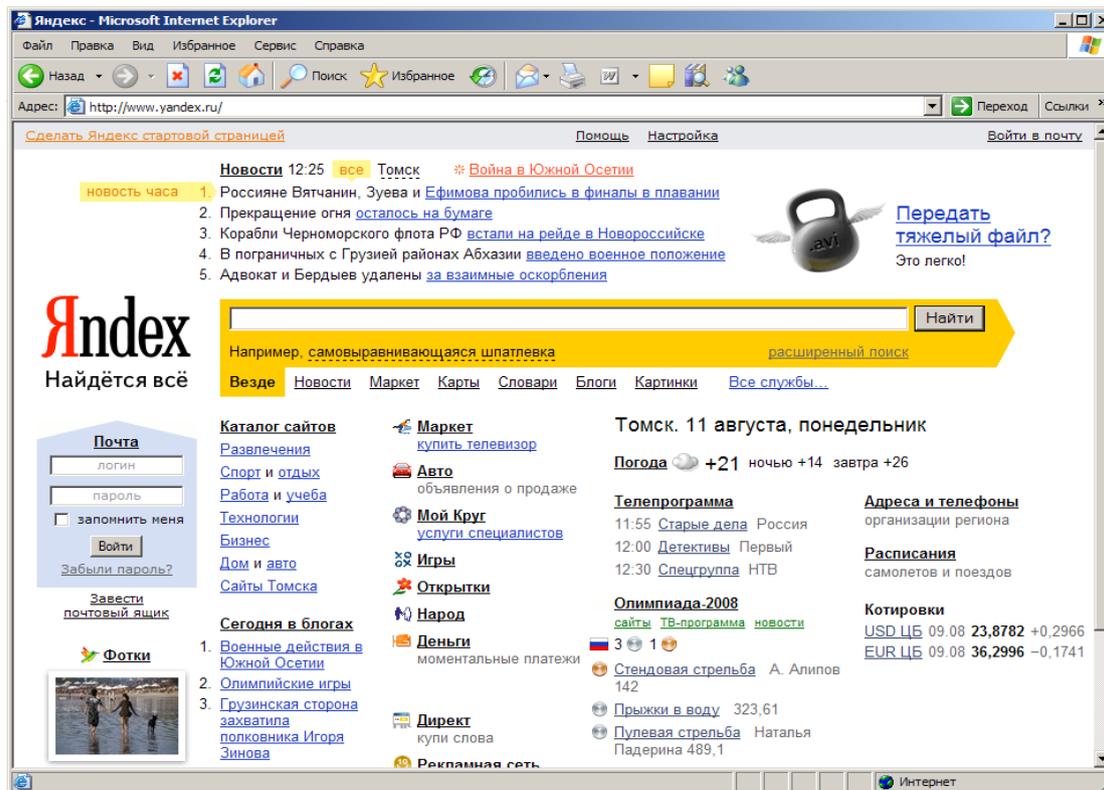


Рис. 11. Домашняя страница поисковой системы Яндекс

По умолчанию слова запроса связываются оператором OR. Под поисковым полем приводится пример формулировки запроса, меняющийся при каждом новом открытии страницы *простого поиска*. Селекторные кнопки снизу позволяют ограничивать поиск следующими областями: «Новости», «Маркет», «Карты», «Словари», «Блоги», «Картинки».

Поиск ключевых слов производится с учетом их морфологии. Если необходимо осуществить поиск по точной словоформе, то перед ним ставится знак «!» без пробела. Система различает слова, набранные строчными и прописными буквами. Поддерживается поиск по фразе, которая заключается в кавычки. Перед ключевым словом допускается постановка знаков «+» и «-», которые заменяют логические операторы AND и NOT соответственно.

Yandex использует собственную систему обозначений логических операторов, которая одновременно позволяет задавать степень близости между ключевыми словами. Логические операторы AND и NOT обозначаются символами «&» и «~» соответственно. Употребление этих символов определяет наличие или отсутствие ключевых слов в отдельном предложении документа.

Для того чтобы подняться от уровня предложения до уровня всего документа, необходимо удвоить символ-оператор. Например, по запросу *библиотеки && архивы* будут найдены документы, содержащие оба слова без учета расстояния между ними, а по запросу *библиотеки & архивы* – документы, содержащие оба слова в пределах одного предложения.

Логический оператор OR вводится с помощью символа «|» и действует в пределах всего текста документа. Возможно также употребление круглых скобок для составления сложных поисковых предписаний.

Символ «/» ограничивает максимальное расстояние между ключевыми словами определенным числом. Например, по запросу *библиотеки /5 архивы* будут найдены документы, в которых расстояние между ключевыми словами не превышает 5 слов. Комбинация символов «/+» позволяет задать расстояние более точно. Так, запросу *библиотеки /+2 архивы* будут релевантны документы, содержащие фразу «библиотеки, областные архивы».

Также поддерживаются следующие специальные операторы для поиска в определенных областях html-документов:

**\$title** – в заголовке;

**\$anchor** – в тексте ссылок;

**#keywords=** – в ключевых словах (поле «keywords»);

**#abstract=** – поиск в описании (поле «META»);

**#image=** – в имени файла изображения;

**#hint=** – в подписи к изображению;

**#link=** – поиск ссылок на заданный URL-адрес;

**#url=** – поиск документов на заданном сайте (странице).

Интерфейс *расширенного поиска* системы Яндекс представляет собой шаблон, состоящий из поля для ввода ключевых слов, их характеристик по месту расположения и употребления, а также свойств найденных страниц, различающихся по языку, дате и формату (рис. 12).

Результаты поиска выдаются в виде заголовка документа, его описания, URL-адреса, а также ссылок на рубрику каталога Internet-ресурсов *List.ru*, в которую попадает данный документ (рис. 13).

Результаты поиска сортируются по степени релевантности документов запросу и выдаются по 10 документов на страницу (рис. 14).

Релевантность документа зависит от ряда факторов, в том числе от частотных характеристик ключевых слов, их близости в тексте документа, а также от веса слова – параметра, который пользователь может задать самостоятельно. Для этого используется символ «:» и определенное число. Например, по запросу «*городские архивы*» библиотеки:3 будут найдены документы, содержащие фразу «городские архивы» и слово «библиотеки», но чем чаще в документе встречается слово «библиотеки», тем ближе он окажется к началу списка результатов.

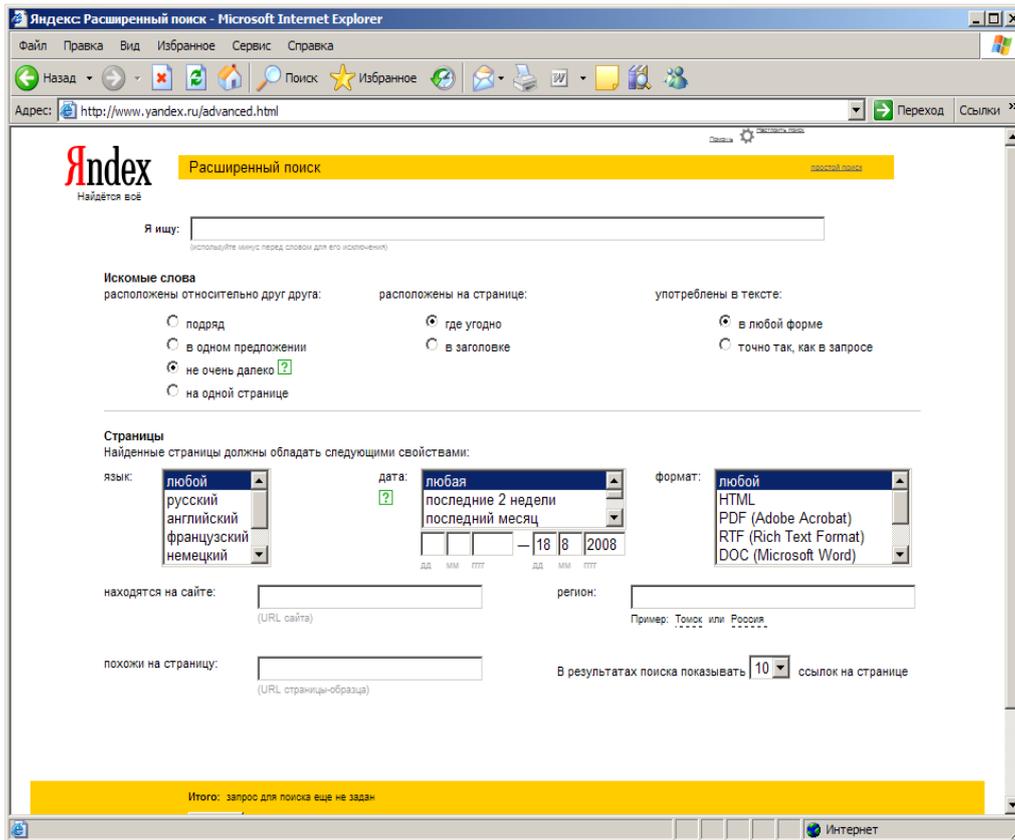


Рис. 12. Интерфейс расширенного поиска системы Яндекс

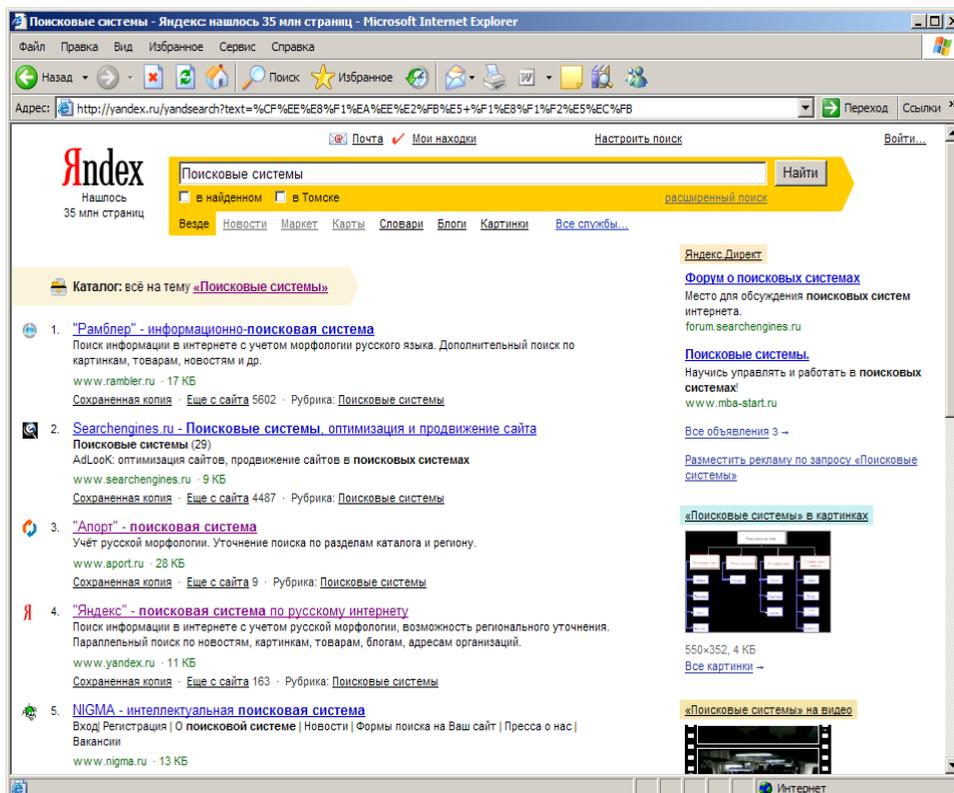


Рис. 13. Результаты поиска системы Яндекс

В числе дополнительных возможностей, предлагаемых пользователям поисковой системы Яндекс, можно назвать следующие: интеграция с каталогом Internet-ресурсов *List.ru*, поиск по новостным лентам ведущих информационных агентств, поиск в электронных магазинах и поиск по российским WAP-ресурсам, а также программа «Региональный Яндекс».

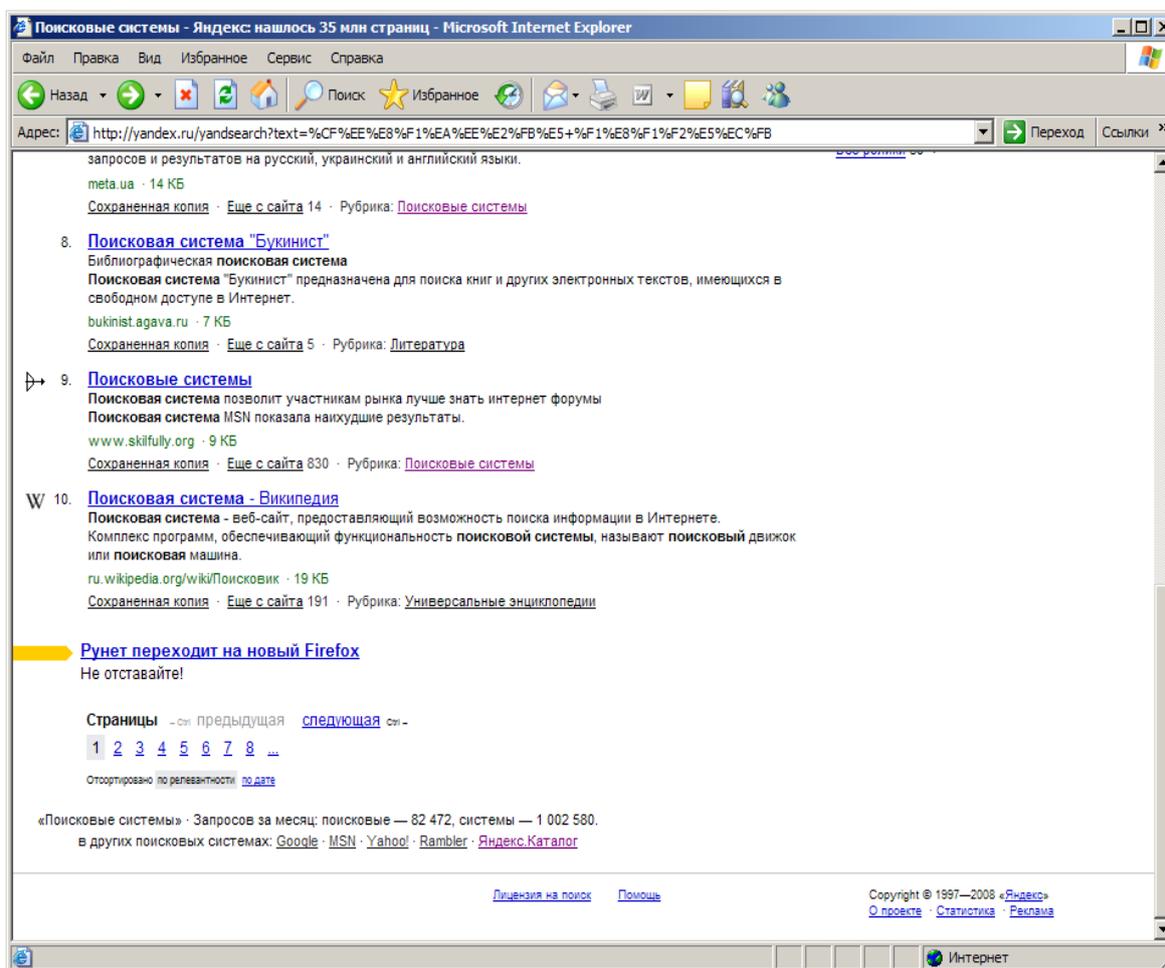


Рис. 14. Формат выдачи результатов поиска системы Яндекс

Поиск по новостным лентам (<http://news.yandex.ru>) ведущих информационных агентств дает пользователям возможность получить список ссылок на полный текст информационного сообщения на сайте того или иного агентства. Также осуществляется доставка последних новостей по интересующей теме по электронной почте.

Раздел «Яндекс.Товары» (<http://товар.yandex.ru>) позволяет искать необходимые товары как в онлайн-овых, так и в обычных магазинах, предоставляющих соответствующую информацию.

Программа «Региональный Яндекс» (<http://www.yandex.ru/regions.html>) была создана благодаря сотрудничеству с региональными каталогами.

Теперь можно ограничивать поиск ресурсами выбранного региона (Астрахань, Брянск, Владивосток, Воронеж, Екатеринбург, Иркутск и проч.).

В разделе «Яндекс.WAP» (<http://wap.yandex.ru>) предлагается два вида поиска: по российским WAP-ресурсам и поиск на территории Москвы. Владелец мобильного телефона, указав свое местоположение (станцию метро или улицу), сможет узнать адреса ближайших станций техпомощи и автозаправки, пунктов обмена валюты и банкоматов, гостиниц, театров, кафе и т. п.

### 8.3.3. Поисковая система Google

Поисковая система *Google* (<http://www.google.com>) была открыта в сентябре 1999 года. На сегодняшний день объем базы составляет более 1 миллиарда документов. Система предлагает пользователю простой и расширенный поисковый интерфейс, а также страницу создания предустановок поиска (рис. 15).

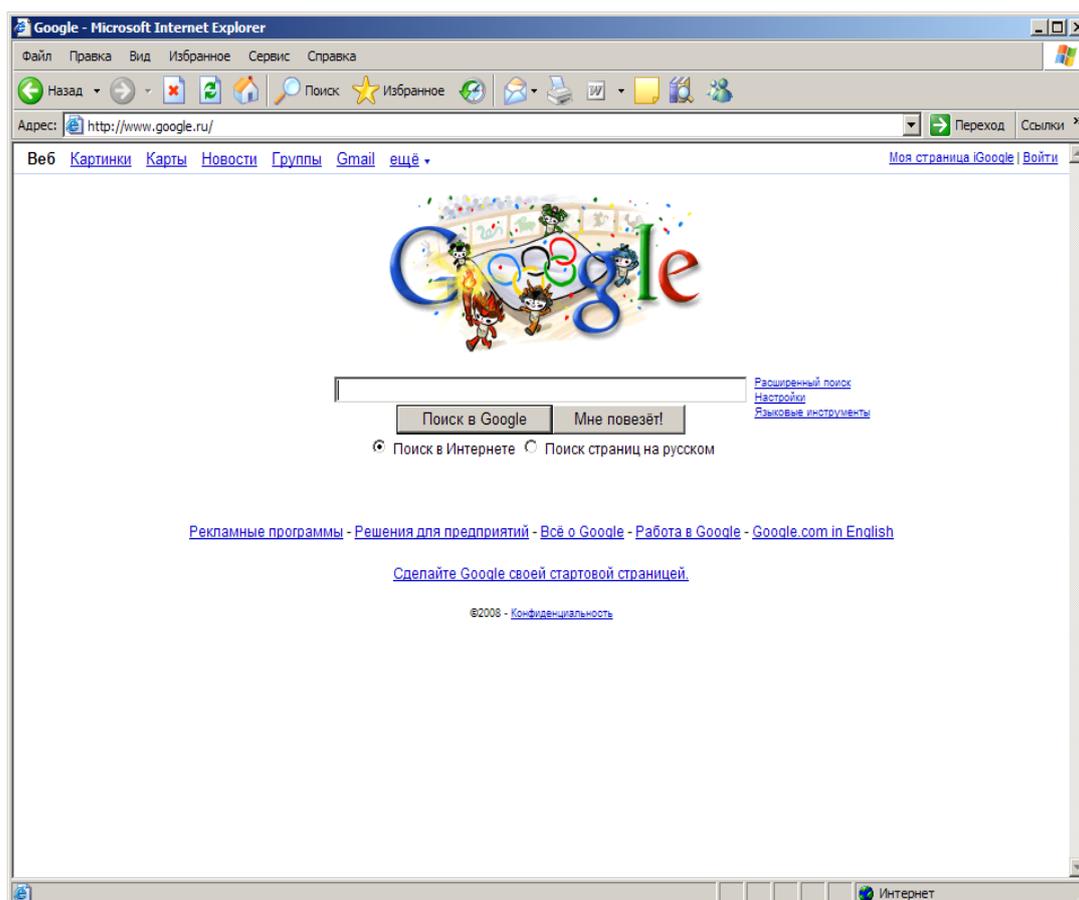


Рис. 15. Домашняя страница поисковой системы Google

Отличительной особенностью Google является технология определения степени релевантности документа путем анализа ссылок других

источников на данный ресурс. Эта технология называется *PageRank™*. Чем больше ссылок на какую-либо Web-страницу имеется на других страницах, тем выше ее рейтинг в базе Google. При выдаче результатов поиска в начале списка оказываются страницы с более высоким рейтингом (при прочих равных составляющих).

Помимо основной базы запрос обрабатывается с использованием таких информационных массивов как БД RealNames и каталог Internet-ресурсов Google Web Directory.

Google предоставляет доступ к своей базе другим поисковым системам, среди которых наиболее известными являются Netscape's Search и Yahoo!.

Основными достоинствами системы являются значительный объем базы, маленький список стоп-слов и возможность получения копии документа из базы Google, если он удален с основного адреса.

Поисковая система Google позволяет осуществлять *простой и расширенный* поиски. При обработке запроса система интерпретирует пробел между словами как логический оператор AND, однако ввод самого оператора не поддерживает. Запрос вводится в поисковое поле. Справа расположены ссылки на страницу «Расширенный поиск» и страницы создания предустановок поиска: «Настройки» и «Языковые инструменты».

Если необходимо провести поиск с использованием стоп-слов, то перед ними проставляется знак «+». Система поддерживает использование логического оператора OR. Оператор NOT заменяется знаком «-» перед словом без пробела. Возможна постановка знаков «+» и «-» перед фразой.

Поддерживается поиск по фразе. Фраза заключается в кавычки. Помимо кавычек Google учитывает следующие знаки препинания, служащие для связи слов: дефисы, косые черты, знаки равенства, апострофы. При поиске слова, связанные этими знаками, воспринимаются как фраза.

Система не поддерживает поиск с учетом морфологии, поиск по части ключевого слова и не различает строчные и прописные буквы.

При составлении поискового выражения можно использовать два специальных оператора. Оператор **link:** дает возможность выявить документы со ссылкой на данный URL. Например, на запрос *link: www.nlr.ru* будут получены документы со ссылками на домашнюю страницу РНБ (Российской национальной библиотеки). Такой запрос нельзя комбинировать с обычными ключевыми словами. Оператор **site:** сужает круг поиска документами с определенного web-сайта. Например, по запросу *site: www.nlr.ru database* будут найдены документы на Web-сайте РНБ, содержащие слово «database».

Интерфейс страницы *расширенного поиска* реализован в виде шаблона, состоящего из фильтров (рис. 16).

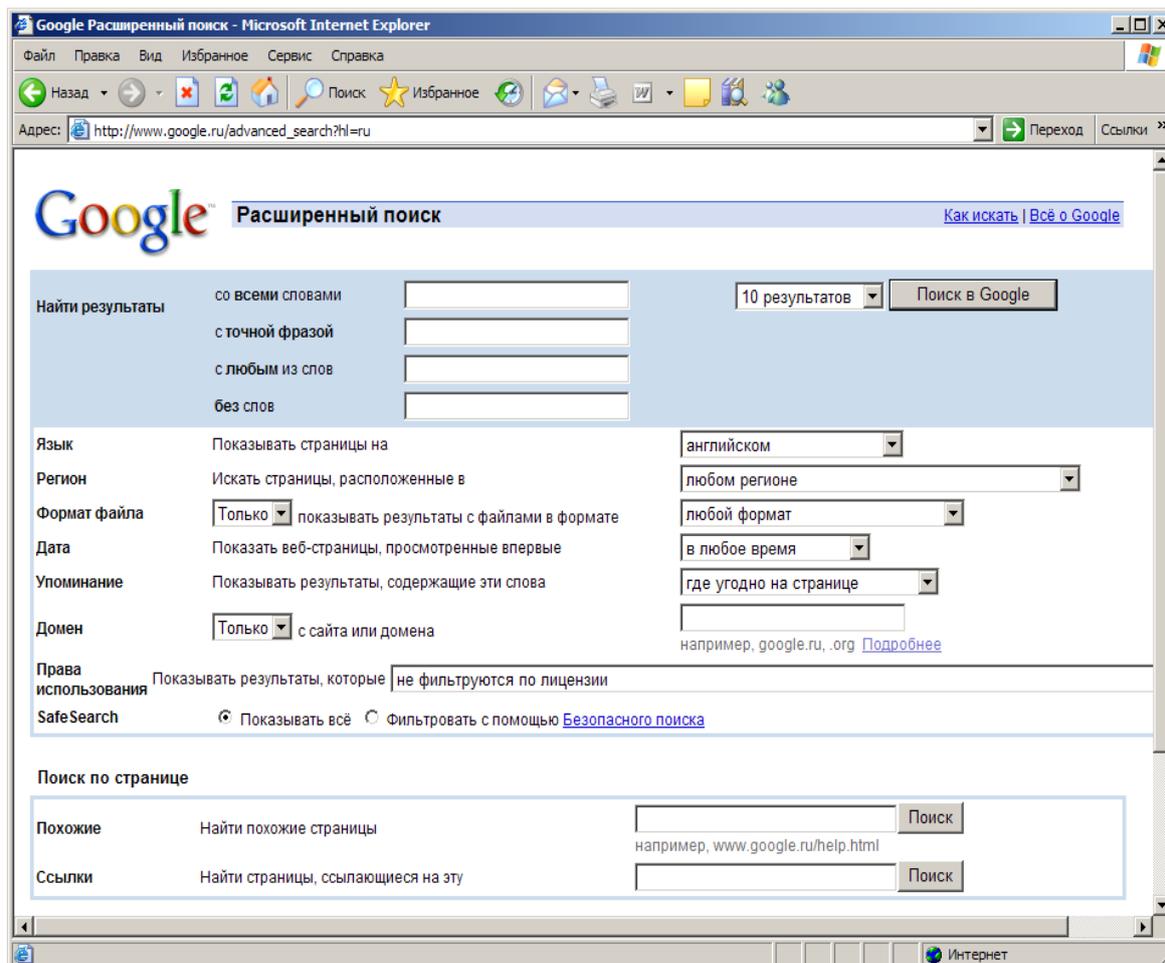


Рис. 16. Интерфейс расширенного поиска системы Google

Поисковая система Google определяет степень релевантности документа путем анализа ссылок других источников на данный ресурс. При сортировке результатов поиска из всех релевантных документов выбираются страницы с более высоким рейтингом и помещаются в начало списка.

Перед списком результатов указывается количество документов, найденных по запросу, и время обработки запроса в базе Google (рис. 17).

Формат вывода результатов поиска состоит из следующих элементов (рис. 18):

- заголовок документа;
- выдержки из текста с выделенными жирным шрифтом словами запроса;
- описание документа, полученное из поля meta;
- ссылка на соответствующий раздел каталога Google Web Directory;
- URL-адрес страницы;
- размер найденного документа в килобайтах;
- ссылка на копию документа в базе Google;

- ссылка для задания поиска документов, наиболее релевантных данному;
- другие страницы сайта, релевантные запросу, если таковые имеются.

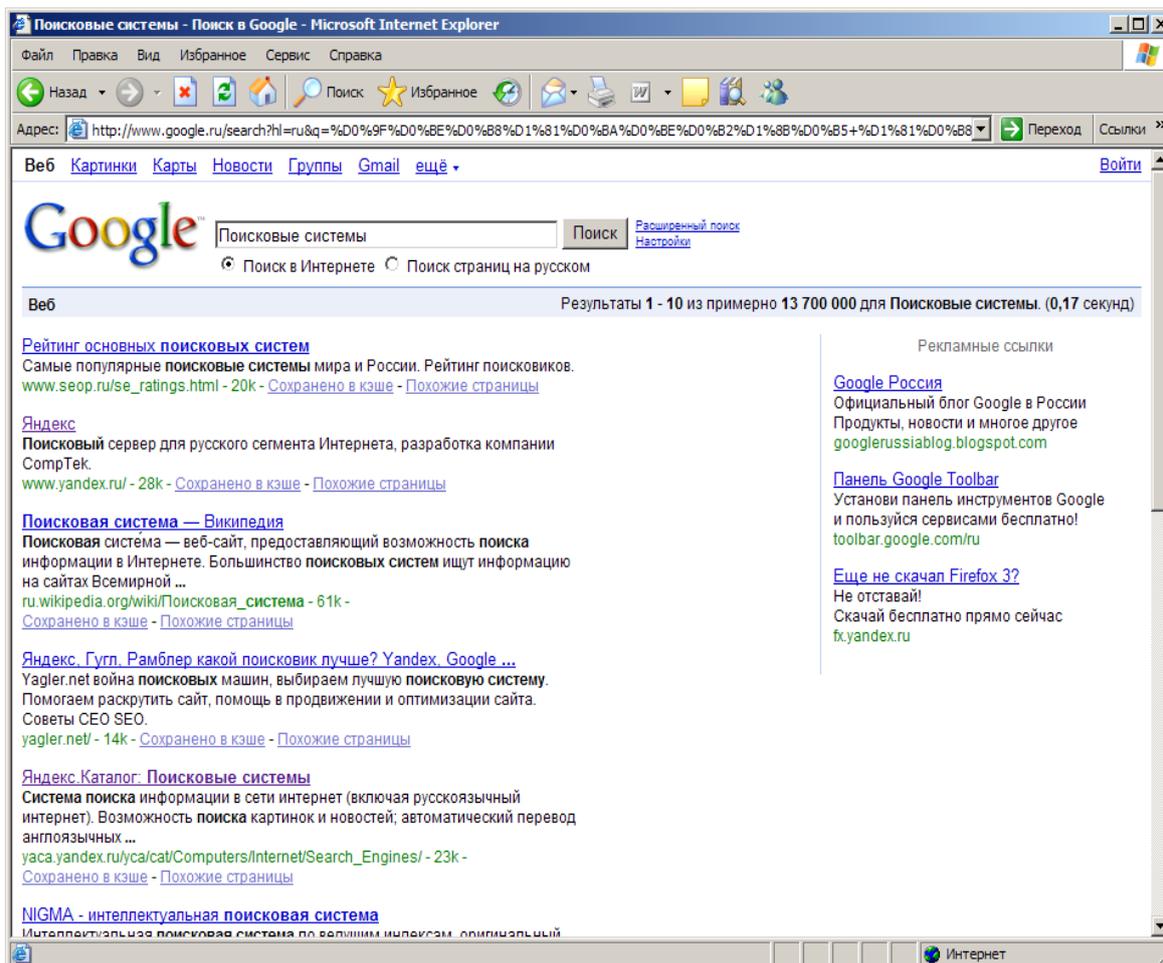


Рис. 17. Результаты поиска системы Google

Оценивая в целом поисковую систему Google, следует отметить, что она предоставляет пользователю разнообразные возможности по настройке интерфейса системы и непосредственно самих поисковых функций. На домашней странице имеется ссылка на страницу создания предустановок поиска: «Настройки» и «Языковые инструменты». Google поддерживает интерфейсы на 43 языках и позволяет открывать каждый найденный документ в новом окне браузера. Пользователь может задать поиск документов одновременно на нескольких языках, регулировать количество результатов поиска, выводимых на одну страницу, а также подключать фильтр для документов, содержащих ненормативную лексику. Созданные предустановки запоминаются браузером и действуют в каждой поисковой сессии, пока не будут изменены.

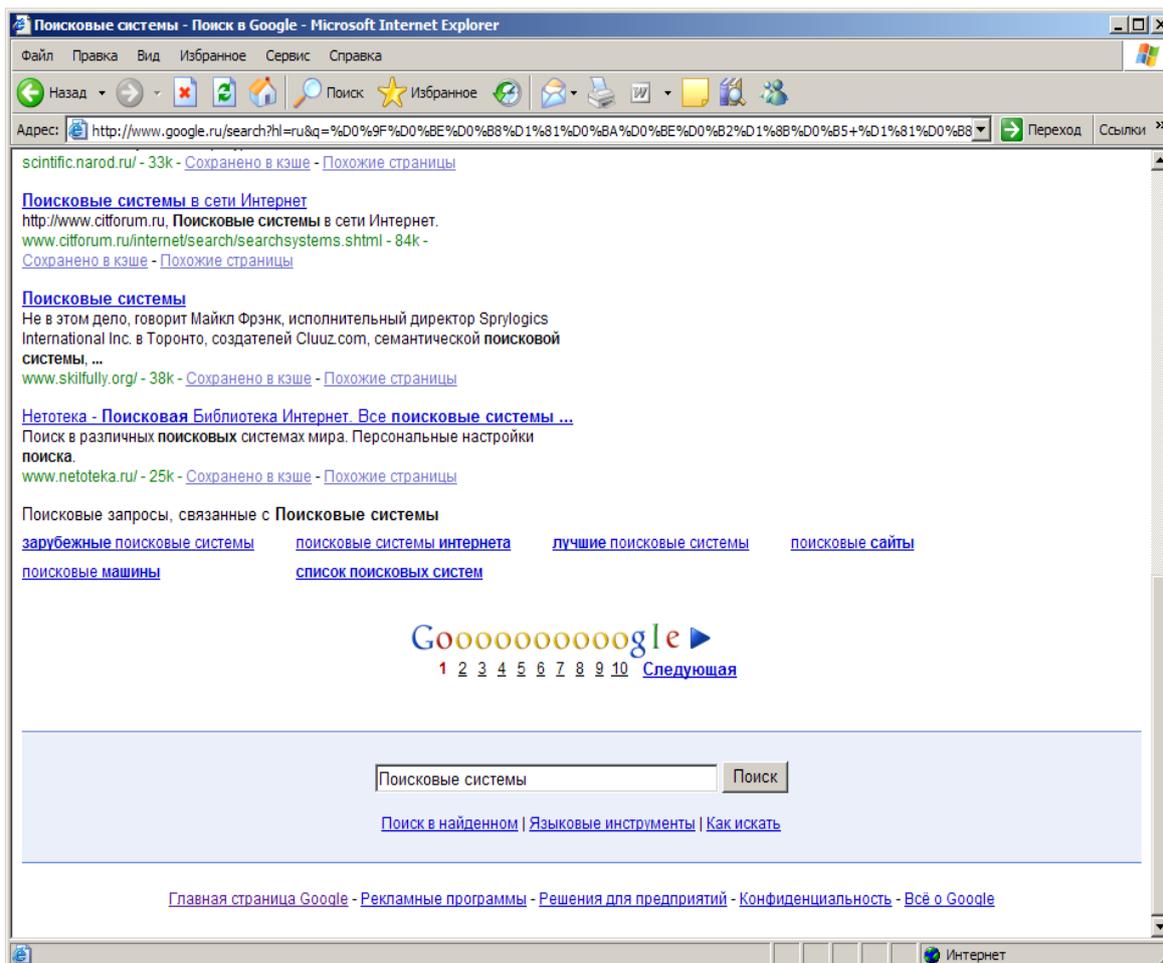


Рис. 18. Формат выдачи результатов поиска системы Google

### 8.3.4. Поисковая система Rambler

Доступ к поисковой системе Rambler (<http://www.rambler.ru>) был открыт в 1996 году. Это одна из первых российских поисковых систем, которая активно развивается. Система использует несколько одновременно работающих программ-роботов и обрабатывает около 5 пользовательских запросов в секунду.

Rambler представляет собой портал, объединивший поисковую систему, рейтинг-классификатор Rambler's Top100, а также ряд бесплатных сервисов и информационных проектов. Ресурсы портала регистрируют ежедневно более 3,5 млн. посещений. Наиболее интересными проектами являются «Rambler-Наука», «Интерактивные карты» и «Словари». Кроме того пользователям предоставляется возможность проведения поиска информации на ftp-серверах (<http://ftpsearch.rambler.ru:8101/>).

Для поиска информации запрос из ключевых слов вводится в поисковое окно, расположенное в верхней части домашней страницы (рис. 19). По умолчанию используется логический оператор «AND».

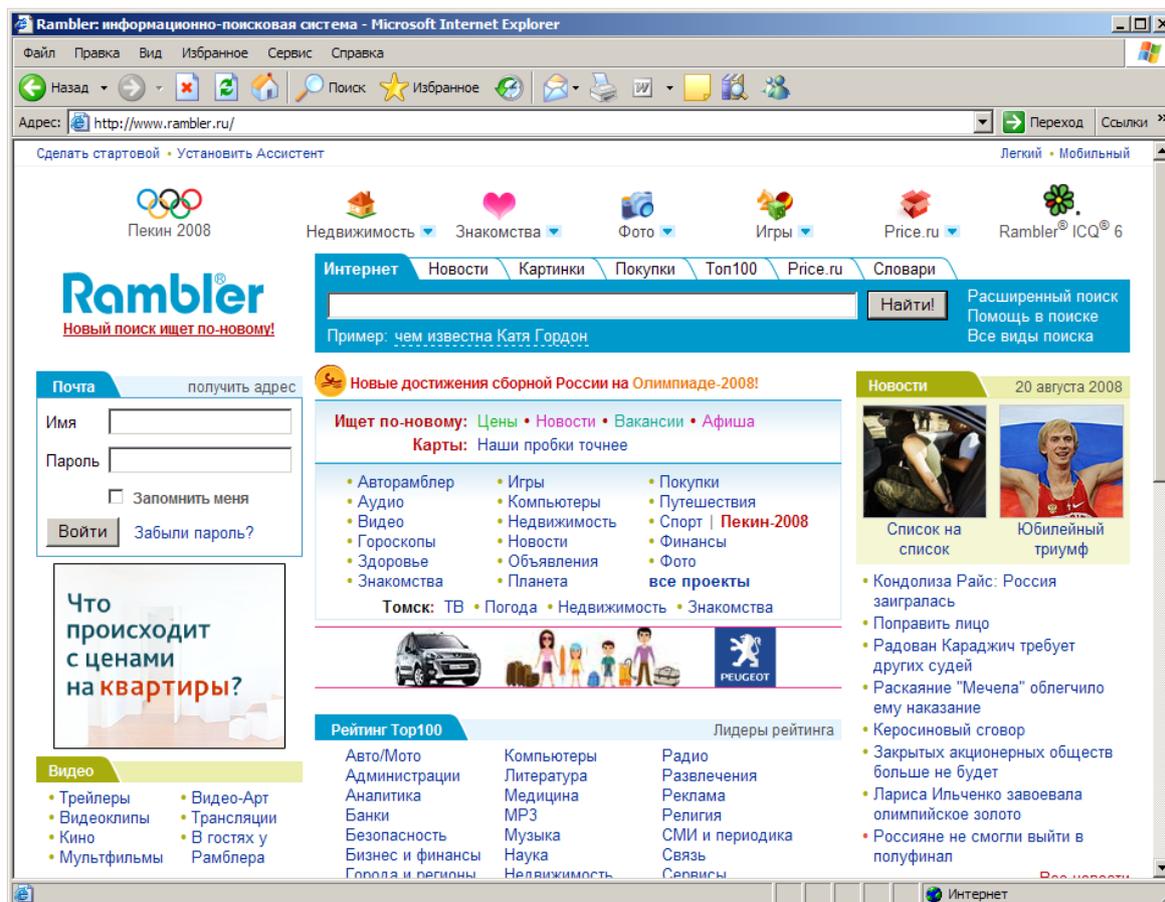


Рис. 19. Домашняя страница поисковой системы Rambler

Система поддерживает составление запроса с использованием логических операторов «AND» («&») и «OR» («|»). Для составления сложного поискового выражения используются круглые скобки, задающие порядок действия операторов. Rambler поддерживает поиск по фразе. Фраза заключается в кавычки. Поиск морфологических форм задается оператором «#», а поиск однокоренных оператором «@». Поддерживается поисковая функция усечения справа и в середине слова. Символ «\*» заменяет любое количество букв; символ «?» заменяет один неизвестный символ.

При составлении запроса можно использовать следующие специальные операторы:

**\$All** – поиск во всех разделах html-документа;

**\$URL** – поиск в URL-адресе html-документа;

**\$Title** – поиск в заголовке html-документа;

**\$Essence** – поиск в аннотации к html-документу.

Интерфейс страницы *расширенного поиска* содержит поле для ввода ключевых слов и шаблон, состоящий из нескольких фильтров для уточнения запроса (рис. 20).

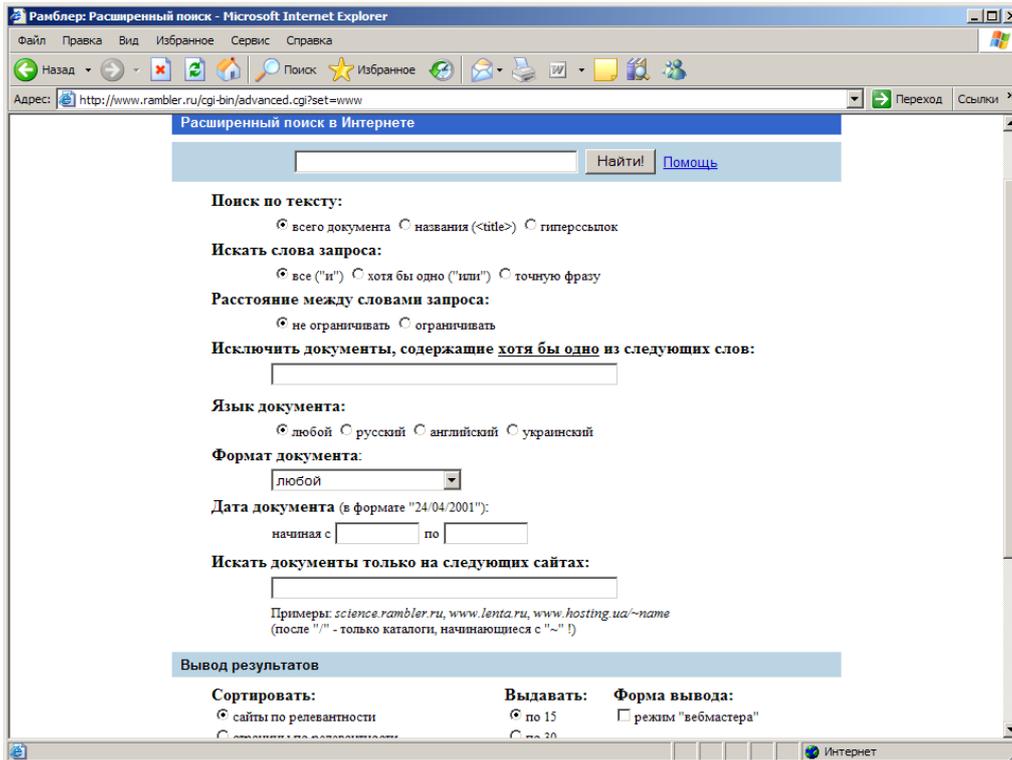


Рис. 20. Интерфейс расширенного поиска системы Rambler

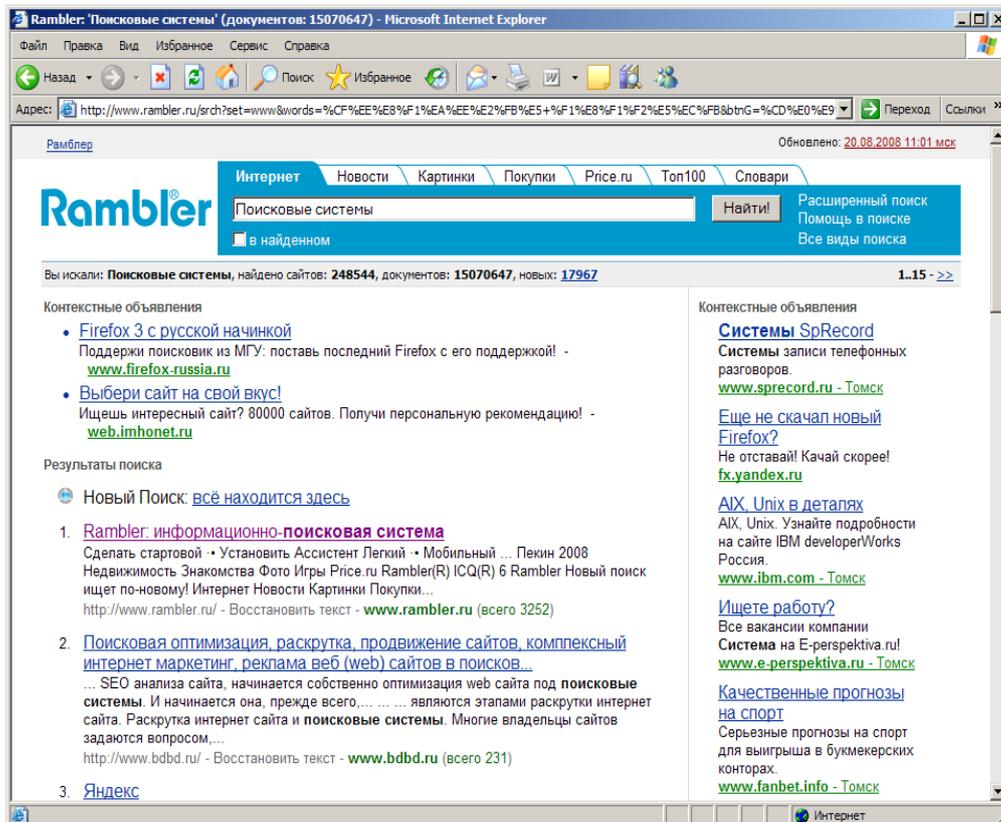


Рис. 21. Результаты поиска системы Rambler

Опция «Искать слова запроса:» содержит селекторные кнопки «все» и «хотя бы одно», заменяющие логические операторы «AND» и «OR». Чтобы исключить документы, содержащие те или иные слова, нужно заполнить поле «Исключить документы, содержащие хотя бы одно из следующих слов:». Система дает возможность ограничить поиск документами, созданными за определенный период времени. Ниже расположена область определения параметров сортировки и выдачи результатов поиска.

По умолчанию найденные документы сортируются по степени релевантности. Каждая страница может содержать от 15 до 50 ссылок на найденные документы (рис. 21).

Полный формат выдачи результатов поиска состоит из следующих элементов: URL-адрес сайта, заголовок документа, выдержка из текста страницы с выделенными жирным шрифтом ключевыми словами, дата создания или последнего обновления документа, ссылки на другие поисковые системы (рис. 22). Кроме того указываются сведения о том, сколько всего найденных документов содержится на данном сайте.

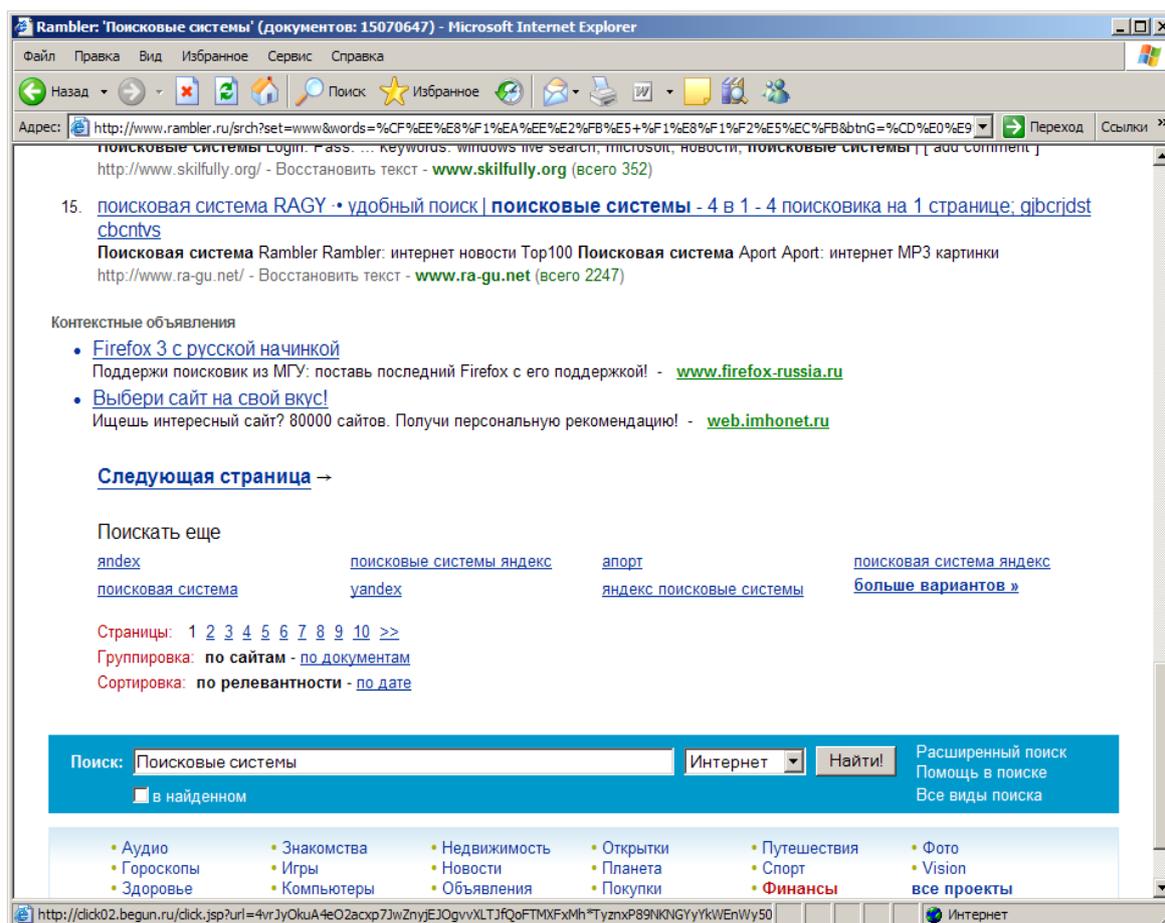


Рис. 22. Формат выдачи результатов поиска системы Rambler

## Методические указания

При изучении этого раздела пособия, содержащего описание принципов работы поисковых систем, необходимо усвоить следующее:

- сервис SE (поисковые системы) – это специальные Web-сайты, на которых пользователь по заданному запросу может получить ссылки на сайты, соответствующие этому запросу;
- поисковая система включает в себя следующие основные компоненты: паук, путешествующий паук, индексатор, база данных, система выдачи результатов, Web-сервер;
- паук – программа, которая скачивает веб-страницы в html-коде для последующей их обработки;
- путешествующий паук – программа, которая выделяет все ссылки, присутствующие на странице, следуя по найденным ссылкам, осуществляет поиск новых документов, еще неизвестных поисковой системе;
- индексатор – программа, которая разбирает страницу на составные части, анализирует (индексирует) их по определенным признакам и формирует index-файл;
- база данных – это хранилище всех index-файлов, полученных поисковой системой в процессе скачивания и анализа Web-страниц;
- система выдачи результатов – занимается ранжированием страниц на соответствие запросу пользователя и определяет порядок выдачи результата поиска;
- Web-сервер – сервер, который осуществляет взаимодействие между пользователем и остальными компонентами поисковой системы;
- архитектура поисковой системы включает в себя две составные части: поисковый робот (паук + путешествующий паук + индексатор), который непрерывно индексирует Web-страницы и заносит их в базу данных в виде index-файлов, и системы выдачи результатов, которая извлекает запрашиваемую пользователем информацию из index-файла;
- для ранжирования страниц в поисковой выдаче используются текстовые критерии, ссылочные критерии и критерии пользовательской оценки;
- релевантность документа – показатель, отражающий соответствие содержания документа конкретному запросу поисковой системы;
- для оценки значимости фрагментов текста применяют формулу Г. Луна;
- для выявления ключевых слов обычно использует статистический частотный анализ по методике В. Пурто;

- в основе ссылочного критерия лежит индекс цитирования, определяемый числом и значимостью ссылок на других сайтах на искомый ресурс;
- ссылочный критерий PageRank – это вероятность пользователя попасть на конкретный документ в зависимости от количества ссылок на него с других документов и от того, насколько вероятно нахождение пользователя на одном из ссылающихся документов и сколько исходящих ссылок содержит этот документ;
- в качестве критерия пользовательской оценки используется коэффициент популярности, определяемый числом пользователей, которые просматривали данную страницу за последние несколько недель;
- формула релевантности документа в целом учитывает релевантность текста документа, релевантность текста с учетом ссылок с других документов и PageRank этого документа;
- основными международными поисковыми системами являются Google, Yahoo и MSN Search, русскоязычного Internet – поисковые системы Яндекс, Google, Rambler.

## ГЛАВА 9. СЕРВИС IRC – ТЕЛЕКОНФЕРЕНЦИИ РЕАЛЬНОГО ВРЕМЕНИ

### 9.1. Назначение сервиса IRC

*IRC (Internet Relay Chat, чат)* – сервисная система, при помощи которой можно общаться через сеть Internet с другими людьми в режиме реального времени. Она была создана в 1988 году финским студентом Ярко Ойкариненом. В русском компьютерном сленге IRC называют «мирк» (калька с названия наиболее популярного клиента – программы *mIRC*).

IRC состоит из серверов. Сервер может быть соединён с другими серверами. Совокупность серверов, соединённых друг с другом, образует сеть IRC. В мире существует множество сетей. Наиболее старые и известные – это IRCNet и EfNet, образовавшейся из расколовшейся первой IRC-сети.

Для работы с IRC необходима специальная программа – клиент. Для Windows-систем клиентом является программа *mIRC*, которая проста в настройке и установке и поэтому 90 % пользователей применяют эту программу.

IRC предоставляет возможность как группового, так и частного общения. Для группового чата в IRC предназначены *каналы*, на которых пользователи могут собираться и вести общение. Общение напоминает общение людей в комнате – кто-нибудь говорит и все его слышат. У каждого канала есть имя, которое отражает общую тематику разговора. Это может быть как интересующая всех тема, так и просто какое-нибудь объединяющее всех название. Если в первом случае разговоры обычно ведутся вокруг указанной тематики, то во втором случае разговоры ведутся на совершенно разнообразные темы.

Кроме обычных пользователей сети IRC есть так называемые *операторы каналов и операторы сети*. Операторы канала – это *модераторы*, которые следят за порядком. Они могут закрыть доступ на канал без приглашения, исключить пользователя с канала, поставить ему *бан* (запрет входа на канал) и многое другое. Оператором канала в классическом случае становится тот, кто первым заходит на канал и тем самым создаёт его. У таких каналов есть постоянные операторы. Для того, чтобы они получили статус оператора при входе на канал используются обычно постоянно сидящие на канале *боты (роботы)*, которые имеют статус оператора и могут его раздавать определённым пользователям.

*IRC-боты* – это программы, которые подключаются к IRC-серверу и могут выполнять специальные операции. Кроме поддержки статуса оператора, боты могут выполнять функции доски объявлений, могут выводить информацию о заходящем на канал пользователе и т. п.

*Операторы IRC-сети* – это администраторы, управляющие работой сети в целом. В их власти исключить пользователя с сервера и запретить ему вход на отдельный сервер или вообще на все сервера сети. Однако, как правило, в IRC-сетях для операторов обычно устанавливаются ограничения на уровне правил сети, сетевого этикета или даже на уровне IRC-сервисов (*IRC-служб*).

*IRC-службы* – это вспомогательные боты в IRC, которые обычно включают в себя:

- NickServ – служба, управляющая пользователями;
- ChanServ – служба, управляющая каналами;
- MemoServ – служба, позволяющая отправлять заметки, когда пользователь не в сети;
- OperServ (RootServ) – служба, позволяющая операторам сети управлять ею;
- Global – служит для оповещения о событиях сервера;
- HelpServ – предоставляет справку по службам IRC.

Также некоторые версии служб содержат:

- BotServ – предоставляет возможность владельцам каналов приглашать на канал ботов;
- HostServ (иногда включён в NickServ) – служба, позволяющая изменять реальный IP пользователя на определённую маску;
- StatServ – служба статистики;
- DevNull – игнорирует все посылаемые данные;
- SeenServ – выводит данные о том, когда в последний раз пользователь появлялся на сервере.

## 9.2. Подключение и настройка mIRC

Подключение к IRC выполняется следующим образом:

1. Для начала необходимо загрузить mIRC (<http://ircinfo.ru/soft/>). Теперь можно приступить к его настройке на примере mIRC 6.31.
2. После инсталляции, во время первого запуска mIRC появится следующее окно (рис. 23). Выбираем категорию «Продолжить».
3. Появится окно для задания параметров пользователя (рис. 24). Обязательно нужно ввести следующие данные:
  - *Ваше имя* – то, что будет отображаться как ваше реальное имя (желательно, его здесь и указать);
  - *Email* – не обязательно указывать ваш реальный e-mail, можно текст [такого@вида];
  - *Ник* – ваш ник (сетевое имя);
  - *Запасной ник* – ник, который будет использоваться, если обычный ник находится online и занят.

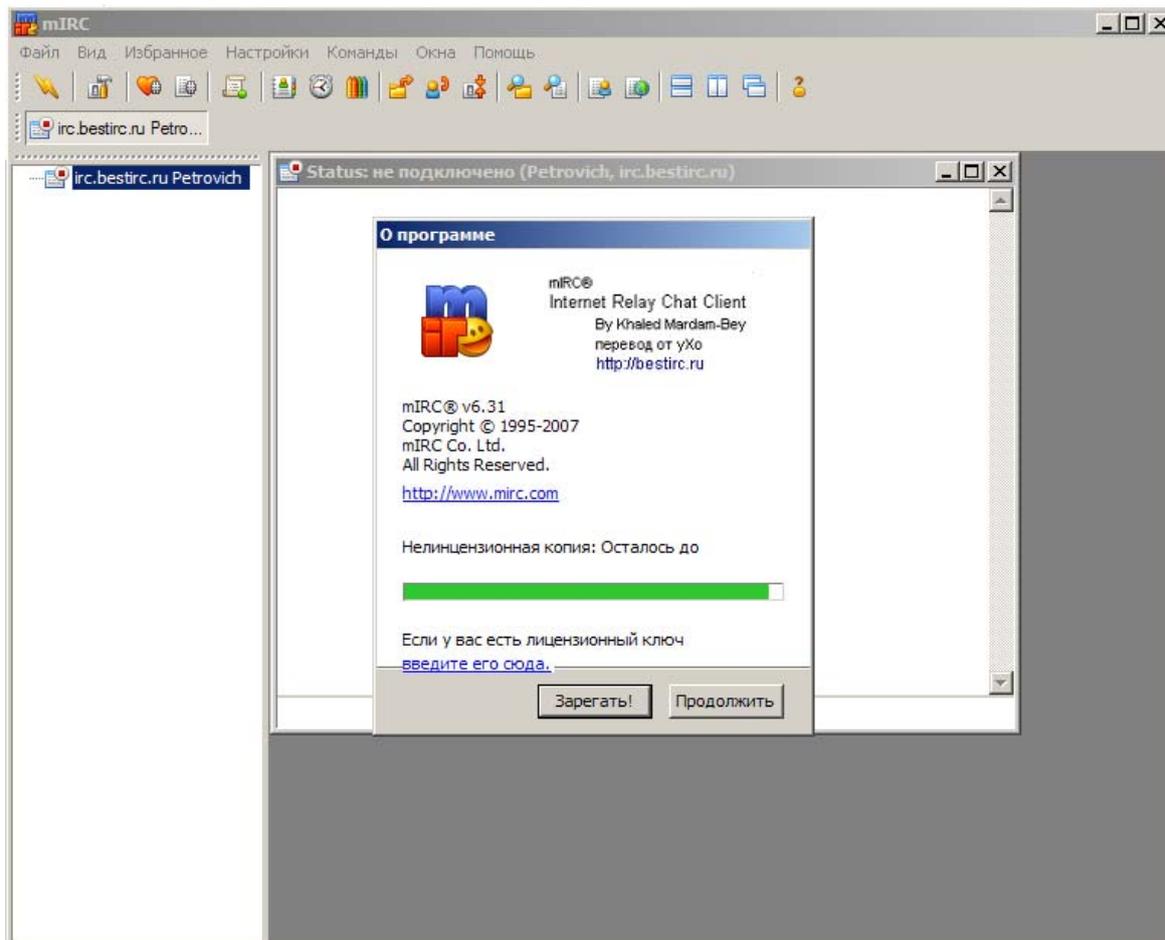


Рис. 23. Окно для подключения к mIRC 6.31

4. Переместимся в меню «Серверы» и добавим наш сервер в список (пункт «Добавить»). При этом необходимо указать следующие сведения:
  - *Описание* – впишем IRCInfo Server;
  - *IRC Сервер* – здесь указываем конкретный адрес ircinfo.ru;
  - *Порт(ы)* – 6667;
  - *Группа* – здесь можно указать сеть, к которой принадлежит сервер – Wenet;
  - *Пароль* – в данном случае пропускаем, в некоторых сетях сюда можно вписать пароль от ника для автоматической идентификации;
5. Сейчас можем соединиться с сервером, для чего нажимаем на значок «Подключиться к серверу» (рис. 26).
6. Для подключения к желаемому каналу нужно ввести команду «/list». При этом откроется окно для доступа к каналам на различные темы (рис. 27).

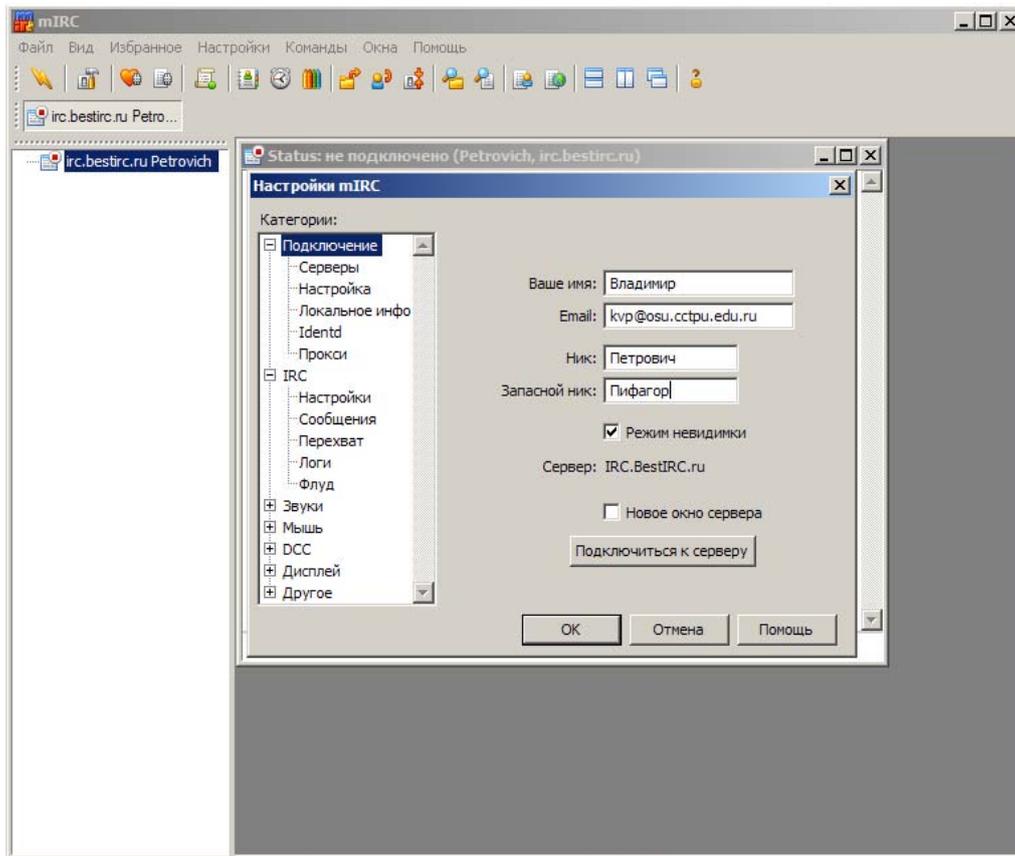


Рис. 24. Окно для задания параметров пользователя

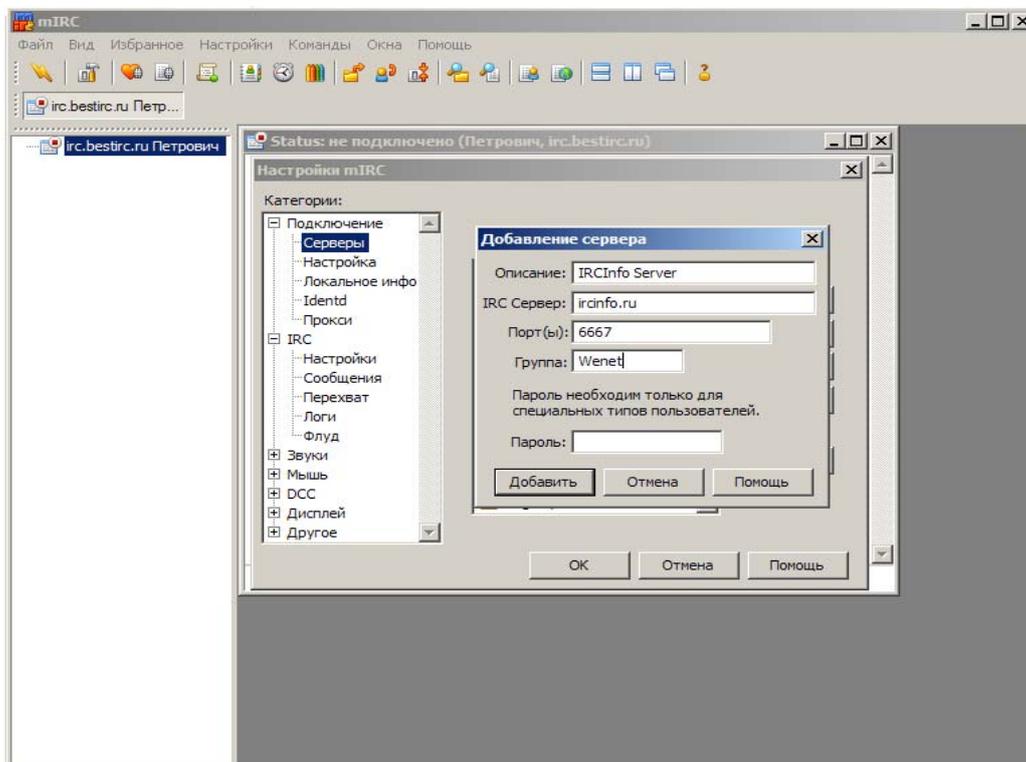


Рис. 25. Окно для добавления сервера

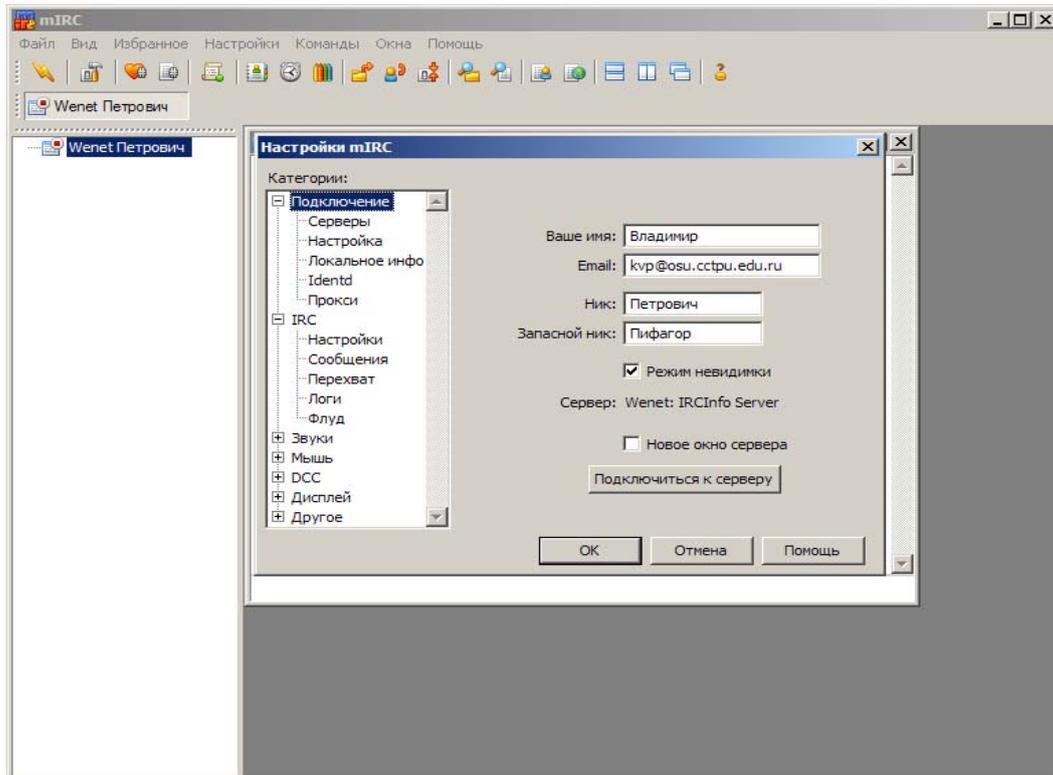


Рис. 26. Окно для подключения к серверу

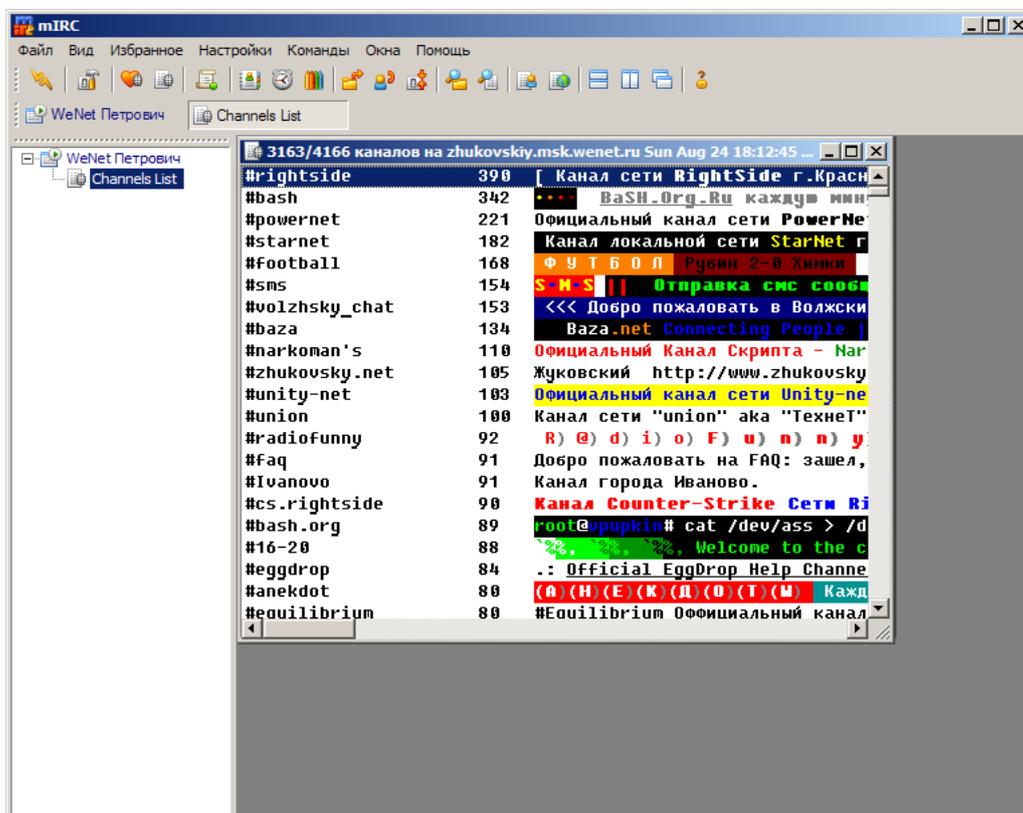


Рис. 27. Окно для доступа к каналам на различные темы

## 9.3. Команды IRC и mIRC

### 9.3.1. Список команд IRC и mIRC

Современные IRC-клиенты, в том числе и mIRC, используют следующий набор команд:

- /nick** – смена ника;
- /identify** – авторизация на NickServ, автоидентификация;
- /whois** – информация о нике;
- /whowas** – информация о нике в оффлайне;
- /mode** – флаги ника;
- /join** – заход на канал;
- /part** – выход с канала;
- /list** – список каналов;
- /invite** – приглашение на канал;
- /msg** – сообщение пользователю/на канал;
- /amsg** – сообщение на все открытые каналы;
- /query** – открытие окна привата;
- /me** – надпись от третьего лица;
- /ame** – надпись от третьего лица на всех каналах;
- /notice** – нотис;
- /kick** – кикнуть пользователя;
- /mode** – флаги каналов, как забанить, запись исключений.

### 9.3.2. Работа с никами

**/nick** *новый\_ник*

Используется для смены ника. В IRC-сетях, в зависимости от используемого в них типа сервера, могут быть различные ограничения на длину ника и использование в нем различных специальных символов, а так же символов национального алфавита.

Подсказка: чтобы добавить в ник пробел, зажмите клавишу <Alt> и на цифровой клавиатуре наберите 0160. Пример: /nick суперник.

**/identify** *пароль*

Авторизовывает ваш ник на NickServ – сервис для работы с никами. Внимание, команда работает не во всех IRC сетях, используйте /msg nickserv identify пароль или /nickserv identify пароль.

Подсказка: после регистрации ника его следует идентифицировать паролем каждый раз при заходе в сеть или настроить автоидентификацию, для mIRC нажмите Alt+R, и первой строкой добавьте:

```
on 1:NOTICE:*IDENTIFY*?:{ if $nick == NickServ /nickserv identify
ПарольНаНик }
```

Пример: /identify F9sk12.

**/whois** *ник*

Покажет информацию о нике, если он находится в онлайн.

Подсказка: если ник зарегистрирован, то дополнительную информацию можно получить командой /ns info ник. Если ник был недавно в онлайн, а потом ушел, то используйте команду /whowas ник.

**/mode** *ник* +|- *флаг*

Использование команды /mode для ника позволяет менять некоторые его настройки и права в текущей сессии. Часть команд доступна только IRC операторам.

Пример пользовательских режимов:

/mode *ник* +i – режим «невидимости» – не присутствовать в выдаче /who, ставится по умолчанию;

/mode *ник* +r – зарегистрированный и проидентифицированный ник;

/mode *ник* +s – получать серверные сообщения и сообщения о килах;

/mode *ник* +H – пользователям без флага +A не будет виден ваш idle-тайм (время бездействия);

/mode *ник* +R – пользователь не будет получать сообщения от ников без +r.

Моды для IRC операторов:

**/mode** *ник* +A – администратор сервера;

**/mode** *ник* +a – администратор сервисов;

**/mode** *ник* +b – получать CHATOPS сообщения;

**/mode** *ник* +c – видеть соединения/отсоединения клиентов;

**/mode** *ник* +e – видеть DCCALLOW трафик;

**/mode** *ник* +f – получать сообщения сервера о флуде;

**/mode** *ник* +k – получать о принудительных отсоединениях клиентов серверами;

**/mode** *ник* +g – получать GLOBOPS сообщения;

**/mode** *ник* +h – присутствовать в ответе на /stats p;

**/mode** *ник* +m – получать предупреждения о спамботах;

**/mode** *ник* +n – получать сообщения маршрутизации;

**/mode** *ник* +o – глобальный оператор;

**/mode** *ник* +O – локальный оператор;

**/mode** *ник* +u – видеть запросы на данные о сервере и сети.

### 9.3.3. Работа с каналами

**/join** *#канал*

Команда необходима, чтобы попасть на канал с известным названием. Если канал не зарегистрирован, то при входе вам дается статус

оператора. Так же, как и на ники, на каналы действует ограничения по количеству символов в названии и использованию специальных символов. Так же, обычно есть ограничение на количество каналов, открытых одним пользователем одновременно (в WeNet – не более 20).

Подсказка:

- чтобы зайти на канал с ключом (паролем), наберите `/join #канал ключ`;
- можно зайти на несколько каналов одновременно, наберите `/join #канал1,#канал2,...`. Количество каналов для захода одной командой ограничено, обратите внимание, названия каналов не отделяются пробелом, только запятой.

`/part [#канал] [причина]`

Используется для выхода с указанного канала или, если он не указан, с открытого в данный момент канала. Причина – не обязательный параметр – будет указана оставшимся на канале посетителям.

Подсказка: как и в команде `/join`, возможен одновременный выход с нескольких каналов: `/part #канал1,#канал2,#канал3`.

Пример: `/part #help` спасибо за помощь.

**/list**

Команда предназначена для вывода списка не скрытых каналов. Обратите внимание, вывод каналов и топиков в популярных IRC-сетях сопровождается получением очень большого входящего трафика.

Подсказка:

- во многих IRC клиентах вы можете сократить выдачу каналов, осуществляя поиск по подстроке, например `/list *love*` выведет каналы, где содержится (в любой части слова) непрерывная цепь символов «love»;
- в `mircc` можно ограничить выдачу каналов по количеству посетителей, например `/list-min 17 – max 20` выведет список, с минимальным числом участников 17 и максимальным –20.

Пример: `/list *help*`

`/invite ник #канал`

Посылает пользователю приглашение зайти на указанный канал, если установлен режим `+i`, является единственно возможным способом попасть на канал. Внимание, массовые приглашения (если поступают жалобы от пользователей) запрещены правилами сети!

Подсказка: если вы оператор канала, то в случае установки режима `+i` можете пригласить себя, для этого наберите `/chanserv invite #канал`.

Пример: `/invite Petja #help`.

### 9.3.4. Отправка текста

**/msg** *ник|#канал текст*

Отправляет сообщение указанному нику без открытия окна привата, может так же использоваться для отправки текста на канал **/msg #канал текст**.

Подсказка: для отправки текста на все открытые каналы используйте команду **/amsg текст** (без указания каналов).

Пример:

- **/msg rrr** Привет!
- **/amsg** Привет

**/query** *ник*

Открывает окно привата с указанным ником.

Пример: **/query helper**

**/me** *текст*

Подсказка: можно отправить текст на все открытые каналы командой **/ame текст**.

Пример: **/me** читает

**/notice** *ник|#канал текст*

Отправка «нотиса» пользователю. Обычно *нотис* используется для сообщения о чем-то важном и его не следует использовать в других случаях.

Подсказка:

- возможна отправка нотиса на канал – **/notice #канал** сообщение, его увидят все пользователи канала и, в большинстве случаев, не будут этим довольны;
- чтобы нотис видели только операторы канала, например, для сигнализации о нарушении правил, используйте **/notice @#канал текст**.

### 9.3.5. Управление каналом

**/kick** *ник [причина]*

Команда используется для выброса с канала нежелательных пользователей, возможно указание причины «кика». Следует помнить, что кик не запрещает возврат пользователя на канал и часто используется как предупредительная мера.

Пример: **/kick Mashka** у нас все в порядке!

**/mode** *#канал +|- флаг [параметр]*

Использование команды **/mode** позволяет операторам управлять каналом и пользователями на канале.

Пример:

/mode #канал +b маска – забанит (то есть запретит вход) пользователя с заданной *маской* (своеобразный адрес пользователя вида nick!ident@host, например aden!aden@staff.wenet.ru);

/mode #канал +t – запретит не операторам менять топик на канале;

/mode #канал –b маска – снимет бан с указанной маски на канале;

/mode #канал +o ник – даст статус оператора пользователю с ником «ник» на канале «#канал»;

/mode #канал +v ник – установит войс на канале.

## Методические указания

В процессе изучения этого раздела пособия необходимо уяснить следующие моменты:

- IRC(чат) – сервисная система, при помощи которой можно общаться через сеть Internet с другими людьми в режиме реального времени;
- совокупность серверов, соединённых друг с другом, образует сеть IRC;
- для группового чата в IRC предназначены каналы, на которых пользователи могут собираться и вести общение;
- кроме обычных пользователей сети IRC есть так называемые операторы каналов и операторы сети;
- операторы канала – это модераторы, которые могут закрыть доступ на канал без приглашения, исключить пользователя с канала, поставить ему бан (запрет входа на канал) и т. п.;
- IRC-боты – это программы, которые подключаются к IRC-серверу и могут выполнять специальные операции: поддержки статуса оператора, функции доски объявлений, выводить информацию о заходящем на канал пользователе и т. п.;
- операторы IRC-сети – это администраторы, управляющие работой сети в целом с помощью IRC-служб;
- IRC-службы – это вспомогательные боты в IRC, которые предназначены для управления пользователями, каналами, оповещения о событиях сервера; предоставлении справки по службам IRC и т. п.;
- для работы с IRC необходима специальная программа – клиент, одной из которых является программа mIRC в среде Windows.
- раздел содержит описание способа подключения и настройки mIRC, список команд IRC и mIRC, правила их применения.

## ГЛАВА 10. СЕРВИС RTVC – ВИДЕОКОНФЕРЕНЦИИ РЕАЛЬНОГО ВРЕМЕНИ

### 10.1. Назначение сервиса RTVC

Видеоконференция – это взаимное общение двух или более лиц по сети с использованием мультимедийных возможностей компьютеров. Компьютер, подключенный к сети, должен быть оснащен видеокамерой, микрофоном и аудиоконками. В зависимости от уровня оборудования, используемого для систем видеоконференцсвязи, различают *персональные, групповые и студийные видеоконференции*.

*Персональные видеоконференции* (рис. 28). Это «начальный уровень» технологий видеоконференцсвязи. Для их создания требуются относительно недорогие программные или программно-аппаратные средства, применяемые на рабочем месте. Такой тип видеоконференцсвязи можно использовать для неформального общения между двумя лицами, обмена интерактивной информацией, пересылки файлов при небольших затратах времени и финансов. В совместной работе с приложениями применяется «доска объявлений» – специализированное приложение, дающее возможность редактировать текстовый или графический документ всем участникам сеанса связи.



Рис. 28. Оборудование персональных видеоконференций

*Групповые видеоконференции* (рис. 29). Применяются для эффективного общения крупных и средних групп пользователей при совместной работе над проектом, для проведения дискуссий и выступлений, на которых участник не может присутствовать лично. Благодаря высокому качеству сигнала можно осуществлять обмен и просмотр документов,

групповую работу с приложениями. Для создания групповых видеоконференции требуются старшие модели видеотерминалов, сервер, обеспечивающий взаимодействие групп пользователей, специализированные программные продукты для рабочих станций и сервера, ISDN-соединение или локальная сеть.

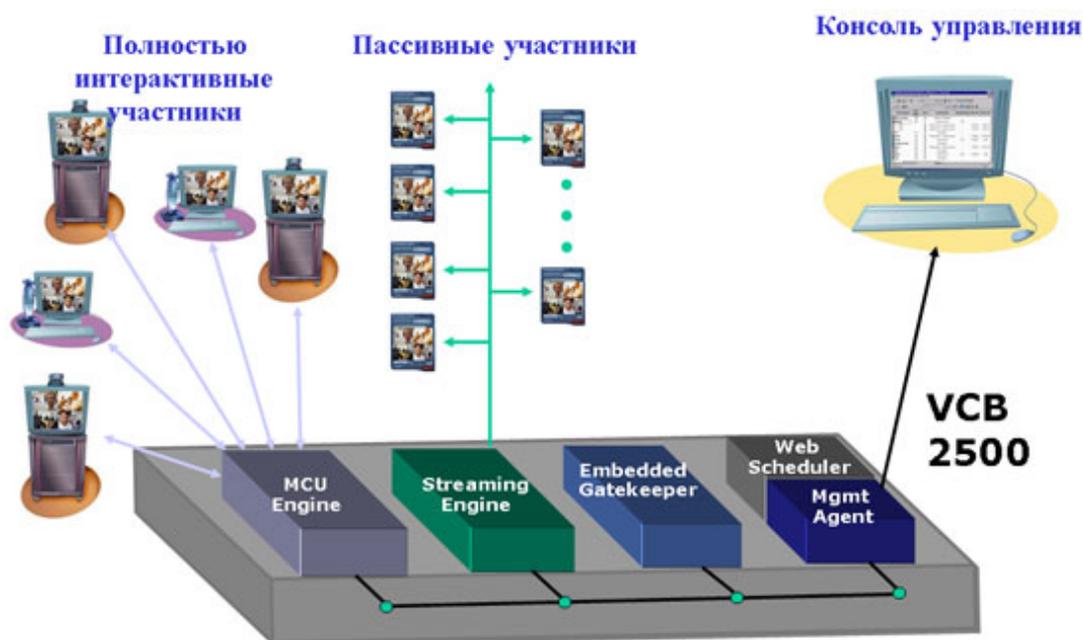


Рис. 29. Оборудование групповых видеоконференций

*Студийные видеоконференции* (рис. 30). Для их создания требуются высококлассное специализированное телеоборудование (студийные камеры, звуковое и контрольное оборудование, мониторы) и максимальная пропускная способность каналов связи (доступ к каналам спутниковой и оптоволоконной связи). Такие видеоконференции используются для решения задач, требующих максимума возможностей с точки зрения организации обработки информации большим числом людей. Для них характерен формальный, жестко регламентированный стиль общения, устанавливаемый ведущим. Типичным примером подобных видеоконференции являются *телемосты*.

По топологии различают два основных типа видеоконференции: *точка-точка* и *многоточечные*. Конференции *точка-точка* наиболее просты. Они подразумевают соединение только двух рабочих станций напрямую, в то время как *многоточечные* видеоконференции дают возможность поддерживать одновременно несколько десятков пользователей или групп пользователей, но при этом требуют дополнительных затрат на установку и поддержку специализированного устройства – сервера управления многоточечными сеансами.

Все терминалы, участвующие в конференции, устанавливают соединение с сервером, который управляет ресурсами видеоконференции, согласовывает возможности терминалов по обработке звука и видео, определяет аудио- и видеопотоки, которые необходимо направлять по многим адресам.

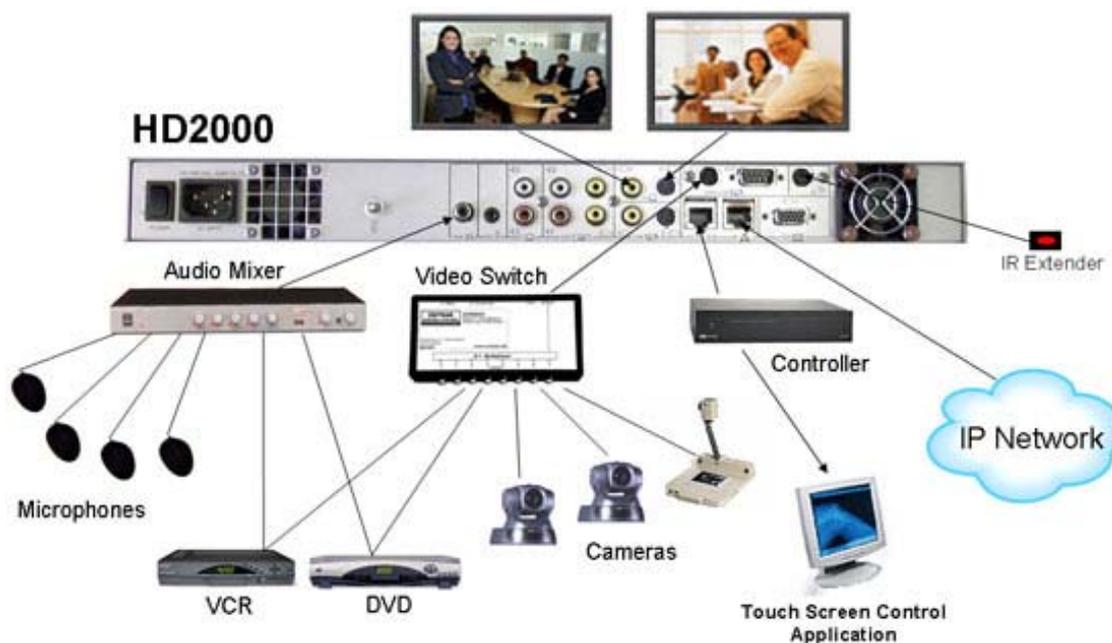


Рис. 30. Оборудование студийных видеоконференций

К примеру, если нужно использовать видеоконференцсвязь с несколькими филиалами только для передачи распоряжений и приема отчетов в реальном режиме времени, то для этой цели вполне подойдет конференция типа «точка-точка» (достаточно обеспечить каждый филиал и головной офис специализированным терминалом), в то время как для организации совещания с участием представителей всех филиалов требуется многоточечная видеоконференция.

Многоточечные сеансы связи могут проводиться в двух основных режимах:

- *активация по голосу* – в этом режиме все участники сеанса видят говорящего, а говорящий видит предыдущего оратора (переключение камеры осуществляется в реальном режиме времени на говорящего);
- *непрерывное присутствие* – на экран каждому участнику поступает изображение от нескольких других участников. При этом экран разделяется на несколько полей – от 2 до 16. Если полей меньше чем участников, то одно из них может работать в режиме «активация по голосу».

И в том и в другом режиме возможен «председательский контроль» выбор активного терминала, подключение и отключение терминалов администратором видеоконференции. При необходимости можно включить автоматический режим администрирования с возможностью в любой момент вмешиваться в этот процесс.

## **10.2. Организация видеоконференций**

### ***10.2.1. Протоколы семейства H.32x***

В 1990 году был одобрен первый международный стандарт в области видеоконференцсвязи – спецификация H.320 для поддержки видеоконференций по ISDN. Затем ITU одобрил еще целую серию рекомендаций, относящихся к видеоконференцсвязи. Эта серия рекомендаций, часто называемая H.32x, помимо H.320, включает в себя стандарты H.321 – H.324, которые предназначены для различных типов сетей.

Во второй половине 90-х годов интенсивное развитие получили IP сети и Интернет. Они превратились в экономичную среду передачи данных и стали практически повсеместными. Однако, в отличие от ISDN, IP сети были плохо приспособлены для передачи аудио и видеопотоков. Стремление использовать сложившуюся структуру IP сетей привело к появлению в 1996 году стандарта H.323 (Visual Telephone Systems and Terminal Equipment for Local Area Networks which Provide a Non-Guaranteed Quality of Service, Видеотелефоны и терминальное оборудование для локальных сетей с негарантированным качеством обслуживания). В 1998 году была одобрена вторая версия этого стандарта H.323 v.2 (Packet-based multimedia communication systems, Мультимедийные системы связи для сетей с коммутацией пакетов), в сентябре 1999 года была одобрена третья версия рекомендаций, 17 ноября 2001 года была одобрена четвертая версия стандарта H.323. Сейчас H.323 – один из важнейших стандартов из этой серии. H.323 – это рекомендации ITU-T для мультимедийных приложений в вычислительных сетях, не обеспечивающих гарантированное качество обслуживания (QoS). Такие сети включают в себя сети пакетной коммутации IP и IPX на базе Ethernet, Fast Ethernet и Token Ring.

Рекомендации H.323 предусматривают:

- управление полосой пропускания;
- возможность организации межсетевых конференций;
- платформенную независимость;
- поддержку многоточечных конференций;
- поддержку многоадресной передачи;
- стандарты для кодеков;
- совместимость и гибкость.

*Управление полосой пропускания.* Передача аудио- и видеoinформации весьма интенсивно нагружает каналы связи, и, если не следить за ростом этой нагрузки, работоспособность критически важных сетевых сервисов может быть нарушена. Поэтому рекомендации H.323 предусматривают управление полосой пропускания. Можно ограничить как число одновременных соединений, так и суммарную полосу пропускания для всех приложений H.323. Эти ограничения помогают сохранить необходимые ресурсы для работы других сетевых приложений. Каждый терминал H.323 может управлять своей полосой пропускания в конкретной сессии конференции.

*Возможность организации межсетевых конференций.* Рекомендации H.323 предлагают средство соединения участников видеоконференции в разнородных сетях (например, IP и ISDN, IP и PSTN).

*Платформенная независимость.* H.323 не привязан ни к каким технологическим решениям, связанным с оборудованием или программным обеспечением. Взаимодействующие между собой приложения могут создаваться на основе разных платформ, с разными операционными системами.

*Поддержка многоточечных конференций.* Рекомендации H.323 позволяют организовывать конференцию с тремя или более участниками. Многоточечные конференции могут проводиться как с использованием центрального MCU (устройства многоточечной конференции), так и без него.

*Поддержка многоадресной передачи.* H.323 поддерживает многоадресную передачу в многоточечной конференции, если сеть поддерживает протокол управления групповой адресацией (такой, как IGMP). При многоадресной передаче один пакет информации отправляется всем необходимым адресатам без лишнего дублирования. Многоадресная передача использует полосу пропускания гораздо более эффективно, поскольку всем адресатам – участникам списка рассылки отправляется ровно один поток.

*Стандарты для кодеков.* H.323 устанавливает стандарты для кодирования и декодирования аудио- и видеопотоков с целью обеспечения совместимости оборудования разных производителей. Вместе с тем стандарт достаточно гибок. Существуют требования, выполнение которых обязательно, и существуют опциональные возможности, в случае использования которых также необходимо строго следовать стандарту. Помимо этого, производитель может включать в мультимедийные продукты и приложения дополнительные возможности, если они не противоречат обязательным и опциональным требованиям стандарта.

*Совместимость.* Участники конференции хотят общаться друг с другом, не заботясь о вопросах совместимости между собой. Рекомен-

дации H323 поддерживают выяснение общих возможностей оборудования конечных пользователей и устанавливают наилучшие из общих для участников конференции протоколов кодирования, вызова и управления.

*Гибкость.* H323 конференция может включать участников, конечное оборудование которых обладает различными возможностями. Например, один из участников может использовать терминал лишь только с аудиовозможностями, в то время как остальные участники конференции могут обладать возможностями передачи/приема также видео и данных.

Основные характеристики протоколов семейства H.32x приведены в табл. 7.

Таблица 7

*Протоколы семейства H.32x*

Рекомендация	H.320	H.321	H.322	H.323 V1/V2	H.324
Год принятия	1990	1995	1995	1996/1998	1996
Сеть	Узкополосная ISDN	Широкополосная ISDN, ATM LAN	Сеть с коммутацией пакетов и гарантированным качеством обслуживания (iso Ethernet)	Сеть с коммутацией пакетов и негарантированным качеством обслуживания (Ethernet)	Аналоговые телефонные сети общего назначения (PSTN или POTS)
Видео	H.261 H.263	H.261 H.263	H.261 H.263	H.261 H.263	H.261 H.263
Аудио	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728 G.723 G.729	G.723
Мультиплексирование	H.221	H.221	H.221	H.225.0	H.223
Управление	H.230 H.242	H.242	H.242 H.230	H.245	H.245
Поддержка многоточечных конференций	H.231 H.243	H.231 H.243	H.231 H.243	H.323	–
Обмен данными	T.120	T.120	T.120	T.120	T.120
Сетевой интерфейс	I.400	AAL I.363 AJM I.361 PHY I.400	I.400 & TCP/IP	TCP/IP	V.34 Модем

### 10.2.2. Базовая архитектура стандарта H.323

В число «объектов» H.323, как они названы в стандарте, включаются терминалы, мультимедиа шлюзы, устройства управления многоточечными конференциями и контроллеры зоны (рис. 31).

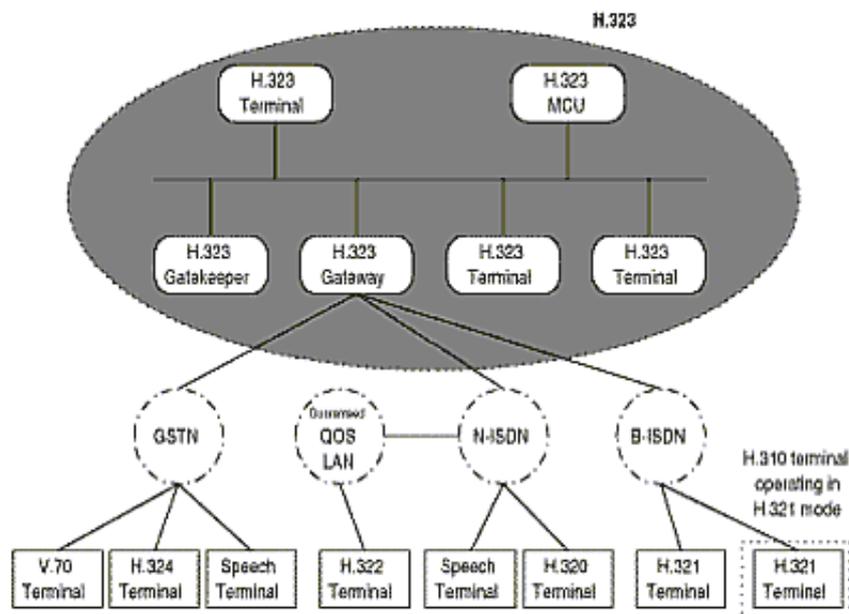


Рис. 31. Базовая архитектура стандарта H.323

*Терминал (Terminal)* – окончательное мультимедийное (голос, видео, данные) устройство, предназначенное для участия в конференции.

*Мультимедиа шлюз (Gateway)* – устройство, предназначенное для преобразования мультимедийной и управляющей информации при сопряжении разнородных сетей.

*Устройство управления многоточечными конференциями (Multipoint Control Unit – MCU)* – предназначено для организации конференций с участием трех и более участников.

*Контроллер зоны (Gatekeeper, Привратник, Конференц-менеджер)* – рекомендуемое, но не обязательное устройство, обеспечивающее сетевое управление и исполняющее роль виртуальной телефонной станции.

Под *терминалом* (рис. 32) стандарт понимает оборудование конечных точек сети, которое позволяет пользователям общаться друг с другом в реальном времени.

Терминалы должны поддерживать протоколы H.245 – согласование параметров соединения, Q.931 – для установления соединения и согласования параметров этого соединения, канал RAS (Registration/ Admission/ Status) взаимодействия с контроллером зоны (Gatekeeper), протокол

RTP/RTCP для работы с потоками аудио и видео пакетов, протокол G.711 для сжатия аудиопотока. Согласно рекомендациям, для терминала H.323 опциональной является поддержка видеокодеков, протокола T.120, и возможностей MCU.

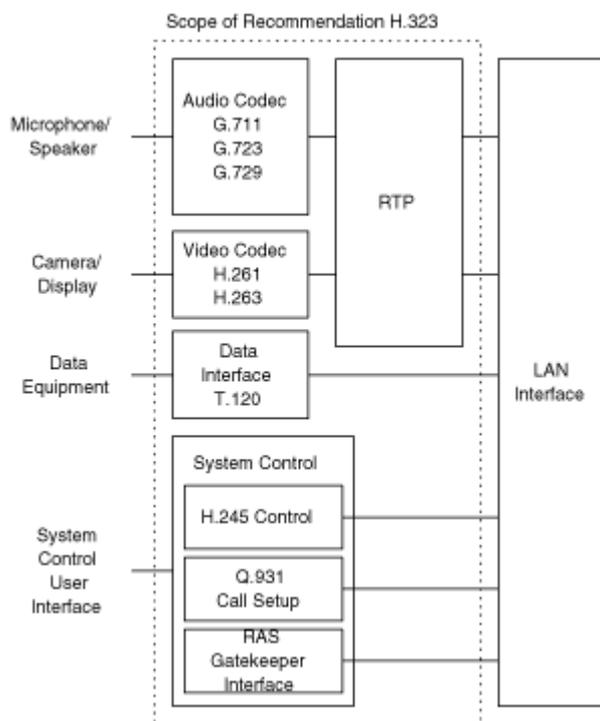


Рис. 32. Структура терминала H.323.

Несмотря на то, что стандарт считает функции видео необязательными, все терминалы с видеовозможностями должны поддерживать кодек H.261, опционально возможна поддержка H.263, который является развитием кодека H.261. Видеокартинка, полученная с помощью кодека H.263 обладает лучшим качеством, поскольку используется полупиксельная технология предсказания движения. Кроме того, используемое кодирование по Хаффману оптимизировано для работы с более низкими скоростями передачи.

*Мультимедиа шлюз* – это опциональный элемент в конференции H.323. Он может выполнять много различных функций. Типичной его функцией являются задача преобразования форматов протоколов передачи (например, H.225.0 и H.221). Обычно мультимедиа шлюзы используются для поддержки взаимодействия между разнородными сетями. На рис. 33 показан шлюз H.323/PSTN.

*Контроллер зоны* (рис. 34) – это рекомендуемое, но не обязательное устройство, обеспечивающее сетевое управление и исполняющее роль виртуальной телефонной станции.

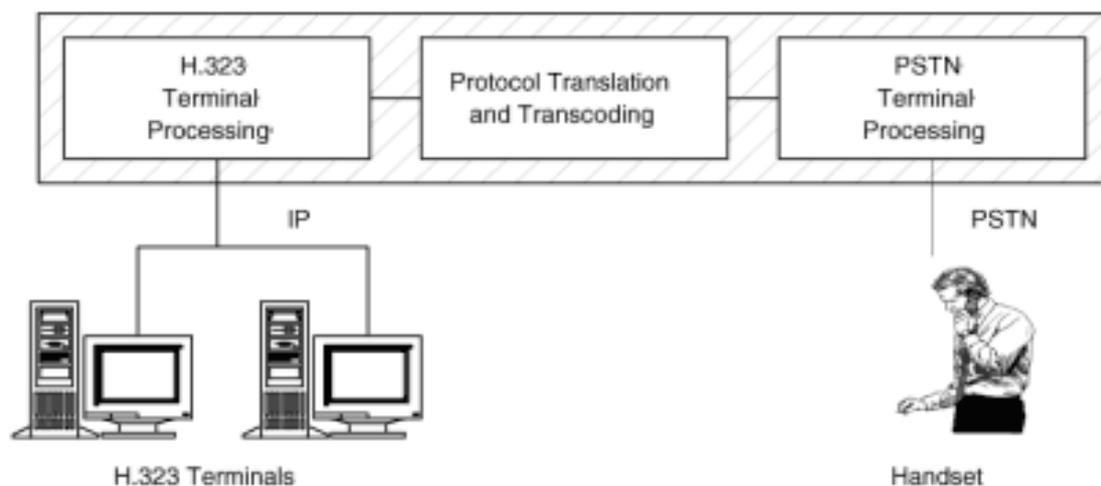


Рис. 33. Мультимедиа шлюз H.323/PSTN

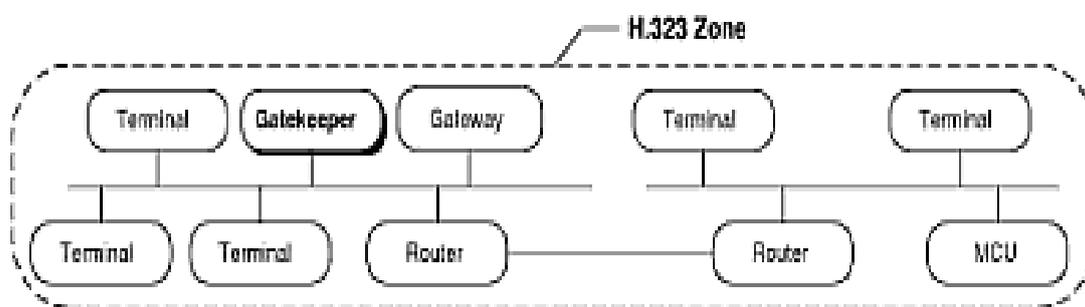


Рис. 34. Контроллер зоны (Gatekeeper)

Основными функциями контроллера зоны являются:

- управление и адресация вызовов;
- обеспечение основными типами обслуживания, такими как телефонный справочник и сервисом, характерным для УАТС (передача и перенаправление вызовов и т. д.);
- управление использованием полосы пропускания приложениями H.323 таким образом, чтобы обеспечить качество обслуживания (QoS).
- управление общим использованием сетевых ресурсов;
- системное администрирование и обеспечение безопасности.

Несмотря на то, что Рекомендации H.323 определяют контроллер зоны как необязательный компонент, без него невозможно воспользоваться мощным и разнообразным спектром услуг, предусмотренных создателями стандарта H.323 для приложений IP-телефонии и мультимедийных телеконференций.

Устройство управления многоточечной конференцией (MCU) – предназначено для поддержки конференции между тремя и более участниками. В этом устройстве должен присутствовать контроллер

Multipoint Controller (MC), и, возможно, процессоры Multipoint Processors (MP). Контроллер MC поддерживает протокол H.245 и предназначен для согласования параметров обработки аудио- и видеопотоков между терминалами. Процессоры занимаются коммутированием, микшированием и обработкой этих потоков.

Конфигурация многоточечной конференции может быть централизованной, децентрализованной (рис. 35), гибридной и смешанной (рис. 36).

*Централизованная* многоточечная конференция требует наличия устройства MCU. Каждый терминал обменивается с MCU потоками аудио, видео, данными и командами управления по схеме «точка-точка». Контроллер MC, используя протокол H.245, определяет возможности каждого терминала. Процессор MP формирует необходимые для каждого терминала мультимедийные потоки и рассылает их. Кроме того, процессор может обеспечивать преобразования потоков от различных кодеков с различными скоростями данных.

*Децентрализованная* многоточечная конференция использует технологию групповой адресации. Участвующие в конференции H.323 терминалы осуществляют многоадресную передачу мультимедиа потока остальным участникам без отправки на MCU. Передача контрольной и управляющей информации осуществляется по схеме «точка-точка» между терминалами и MCU. В этом случае контроль многоточечной рассылки осуществляется контроллером MC.

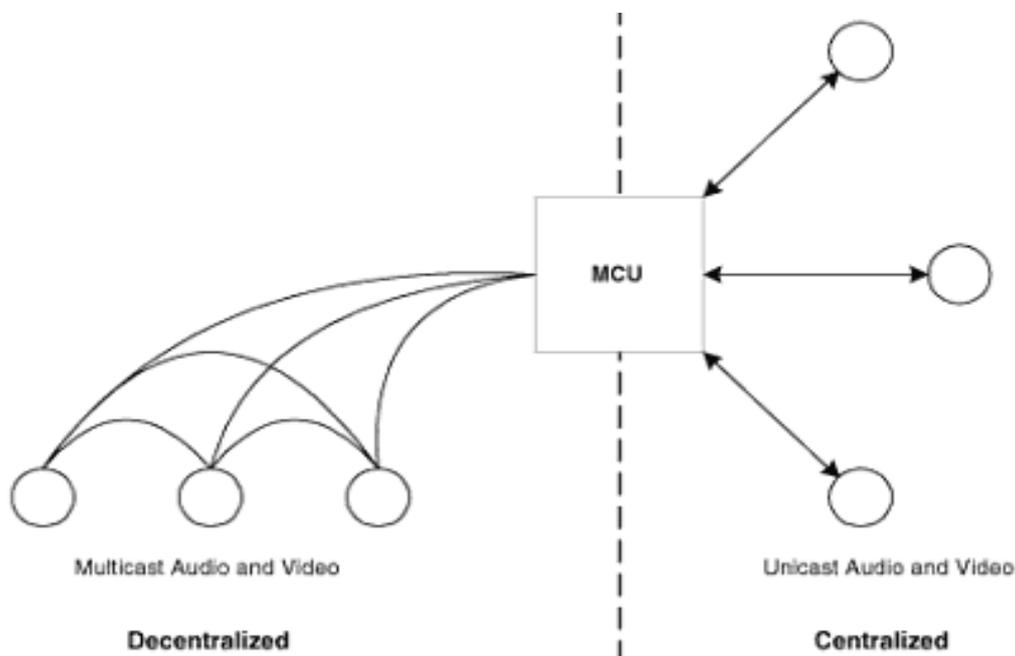
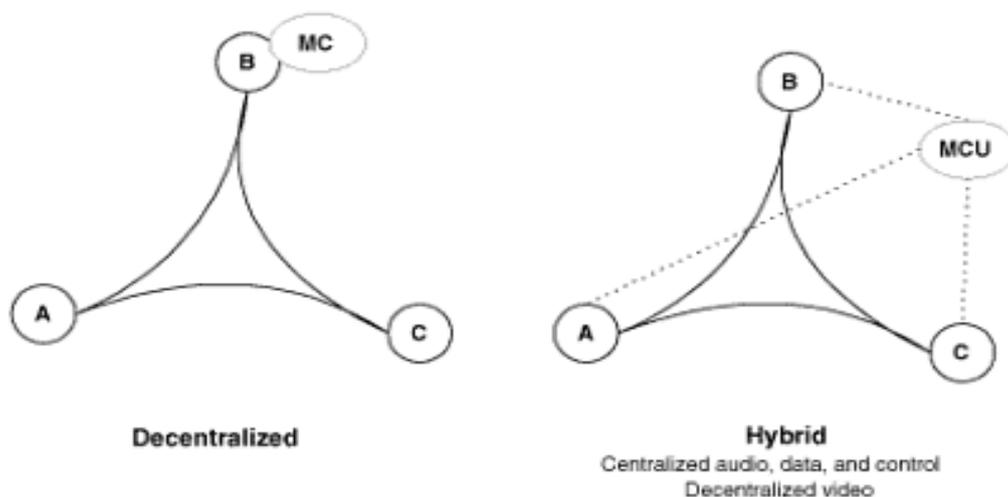


Рис. 35. Схемы централизованной и децентрализованной организации конференции в H.323

*Гибридная* схема организации конференцсвязи является комбинацией двух предыдущих. Участвующие в конференции H.323 терминалы осуществляют многоадресную передачу только аудио- или только видеопотока остальным участникам без посылки на MCU. Передача остальных потоков осуществляется по схеме «точка-точка» между терминалами и MCU. В этом случае задействуются как контроллер, так и процессор MCU.



*Рис. 36. Схемы децентрализованной и смешанной организаций конференции в H.323*

В смешанной схеме организации конференцсвязи одна группа терминалов может работать по централизованной схеме, а другая группа – по децентрализованной.

### **10.2.3. Тенденции развития рекомендаций H.323**

Во второй версии H.323 v.2 рекомендаций были устранены недостатки предыдущей версии. Были усовершенствованы существующие протоколы: Q.931, H.245 и H.225, а также введен ряд новых. Основные преимущества новой версии стандарта заключаются в добавлении функций безопасности, установки быстрого вызова, некоторых дополнительных сервисов и интеграции протоколов H.323 и T.120.

Функции безопасности (H.235) включают в себя обеспечение аутентификации (механизм, который подтверждает то, что участники конференции именно те, за которых они себя выдают), целостности (механизм, подтверждающий то, что переданные пакеты не были искажены), криптографическую защиту передаваемой информации от несанкционированного доступа.

Функция Fast Call Setup решает имевшуюся в первой версии проблему, когда после прохождения звонка одному абоненту другому могла быть задержка в прохождении аудио и видеопотоков.

Протокол Т.120 был интегрирован и в первую версию рекомендаций Н.323, однако сценарии установки звонка были довольно сложны. Во второй версии рекомендаций Н.323 эта проблема решается следующим образом: стандарт требует, чтобы оборудование конечных пользователей, поддерживающее одновременно и Т.120, и Н.323, управлялось звонками по Н.323. Более того, согласно второй версии рекомендаций Т.120 является опциональной частью конференции Н.323 и возможности действий по Т.120 отдаются на усмотрение каждого устройства в Н.323 конференции по отдельности.

В третьей версии Н.323 v.3 рекомендаций было введено несколько новых возможностей. Прежде всего они касаются дополнений к основному документу и рекомендациям Н.225.0, внося усовершенствования в архитектуру стандарта. Среди них можно выделить:

- более эффективное использование ранее установленных сигнальных соединений, в частности, между мультимедиа шлюзом и контроллером зоны;
- возможность переадресации вызова при установленном соединении;
- повышено удобство получения информации об абонентах (Caller ID);
- введена сигнальная информация, которая включает в себя информацию о языке абонента, что расширяет возможности обработки вызова;
- предложен механизм, облегчающий добавление новых кодеков;
- механизм сигнализации может теперь использовать UDP транспорт, вместо TCP, что существенно для конференций с большим числом участников;
- введено понятие упрощенного терминала (Simple Endpoint Type – SET), которые могут поддерживать только незначительную часть рекомендаций Н. 323, тем не менее обеспечивая проведение аудиосвязи с другими Н.323 терминалами;
- введена возможность SNMP – управления оборудованием видеоконференцсвязи;
- информационная база управления (MIB) описывается документом Н.341.

Четвертая версия рекомендаций Н.323 v.4 принята 17 ноября 2000 года. Туда внесено много изменений с целью повышения надежности, мобильности и гибкости систем видеоконференций. Новые возможности, касающиеся мультимедиа шлюзов и устройств многоточечной конференции, направлены на повышение качества организации и проведения конференции с большим числом участников. Перечислим некоторые из нововведений:

- новые механизмы повышения устойчивости работы Н.323 конференции;

- декомпозиция структуры мультимедиа шлюза с целью отделения модуля управления от исполнительных устройств;
- возможность мультиплексирования аудио и видео в одном RTP потоке;
- модификация процесса регистрации на контроллере зоны с целью облегчить регистрацию большого числа участников конференции;
- совершенствование механизмов распределения нагрузки и повышения устойчивости работы контроллеров зоны;
- для терминалов H.323 предусматриваются способы выделения реально необходимой полосы пропускания как для обычной, так и для групповой адресации.

### **Методические указания**

Этот раздел пособия необходим для понимания общих принципов организации видеоконференций. Он содержит следующие ключевые моменты:

- видеоконференция – это взаимное общение двух или более лиц по сети с использованием мультимедийных возможностей компьютеров;
- в зависимости от уровня оборудования, используемого для систем видеоконференцсвязи, различают персональные, групповые и студийные видеоконференции;
- персональные видеоконференции – это общение между двумя лицами для обмена интерактивной информацией или пересылки файлов;
- групповые видеоконференции – применяются для эффективного общения крупных и средних групп пользователей при совместной работе над проектом, для проведения дискуссий и выступлений, на которых участник не может присутствовать лично;
- студийные видеоконференции (телемости) – для их создания требуются высококлассное специализированное телеоборудование и высокоскоростные каналы связи;
- по топологии различают два основных типа видеоконференции: точка-точка и многоточечные;
- конференции точка-точка подразумевают соединение только двух рабочих станций напрямую;
- многоточечные сеансы связи могут проводиться в двух основных режимах: активация по голосу и непрерывное присутствие;
- активация по голосу – в этом режиме все участники сеанса видят говорящего, а говорящий видит предыдущего оратора;
- непрерывное присутствие – на экран каждому участнику поступает изображение от нескольких других участников;

- для организации видеоконференций разработаны протоколы семейства H.32x;
- базовым протоколом семейства H.32x является стандарт H.323, в котором предусмотрено управление полосой пропускания, возможность организации межсетевых конференций, платформенная независимость,
- поддержка многоточечных конференций и т. п.;
- в состав базовой архитектуры стандарта H.323 включены терминалы, мультимедиа шлюзы, устройства управления многоточечными конференциями и контроллеры зоны;
- терминал – оконечное мультимедийное (голос, видео, данные) устройство, предназначенное для участия в конференции;
- мультимедиа шлюз – устройство, предназначенное для преобразования мультимедийной и управляющей информации при сопряжении разнородных сетей;
- устройство управления многоточечными конференциями – предназначено для организации конференций с участием трех и более пользователей;
- контроллер зоны – устройство, обеспечивающее сетевое управление и исполняющее роль виртуальной телефонной станции;
- конфигурация многоточечной видеоконференции может быть централизованной, децентрализованной, гибридной и смешанной;
- стандарт H.323 непрерывно развивается, появились его новые версии: H.323 v.2, H.323 v.3, H.323 v.4.

## ЛИТЕРАТУРА

1. Сетевые технологии. Е-Публиш. 2008.  
<http://net.e-publish.ru/p214aa1.html>
2. В.П. Комагоров. Архитектура сетей и систем телекоммуникаций./ Учебное пособие. – Томск: Томский политехнический университет, 2008. – 154 с.
3. Семенов Ю.А. Телекоммуникационные технологии (v3.1, 19 марта 2008 года). <http://www.citforum.ru/nets/semenov/>
4. Usenet. Материал из Википедии – свободной энциклопедии, 2008.  
<http://ru.wikipedia.org/wiki/Usenet/>
5. Храмцов П. Usenet – система телеконференций Internet, 1996.  
<http://www.lib.ru/LABIRINT/usenet.htm>
6. Приходько В.И. История развития поисковых систем, 2008.  
[http://www.el-book.info/ob\\_poisk\\_history.html](http://www.el-book.info/ob_poisk_history.html)
7. Людкевич С., Есипов Е. Основные факторы, влияющие на релевантность сайта для поисковых систем. «Промо. Текарт», 14.11.2003. [http://www.seop.ru/sites\\_relevant.html](http://www.seop.ru/sites_relevant.html)
8. Поисковые системы русскоязычного Интернет. Корпоративный сайт компании «Корпоративные системы Интернет», 2008.  
<http://corpsite.ru/Encyclopedia/Internet/iService/search.aspx>
9. World Wide Web – Стратегия эффективного поиска: Справочник для библиотек [отдельные главы]. I .УНИВЕРСАЛЬНЫЕ ПОИСКОВЫЕ СИСТЕМЫ. Российская национальная библиотека, 2005.  
<http://vss.nlr.ru/publications/ips.html>
10. IRC. Материал из Википедии – свободной энциклопедии, 18.08.2008. <http://ru.wikipedia.org/wiki/IRC>
11. Команды IRC и mIRC. Автор: IRC Info.RU, 18.08.2008.  
<http://ircinfo.ru/articles/irc-commands.asp>
12. Видеоконференции: вопросы и ответы (FAQ). «Стэл – компьютерные Системы», 2007 г. <http://videoconference.com.ru/q1.php>
13. Видеоконференция Polysom, 2008.  
<http://www.videosvyaz.ru/index.htm>
14. Виноградов М., Метлицкий В. «Стэл – Компьютерные Системы», 2008 г. [http://www.stel.ru/videoconference/tech\\_vc/podrobno/h323/](http://www.stel.ru/videoconference/tech_vc/podrobno/h323/)

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	3
ГЛАВА 1. ПРОТОКОЛЫ И СЕРВИСЫ INTERNET .....	4
1.1. Понятие сервиса Internet .....	4
1.2. Протоколы сервисов Internet .....	6
Методические указания .....	7
ГЛАВА 2. СЕРВИС DNS – СИСТЕМА ДОМЕННЫХ ИМЕН .....	8
2.1. Назначение сервиса DNS .....	8
2.2. Организация пространства имен .....	9
2.3. База данных сервера DNS .....	13
2.4. Разрешение имен .....	14
2.5. Порядок регистрации доменных имен .....	14
Методические указания .....	15
ГЛАВА 3. СЕРВИС TELNET – УПРАВЛЕНИЕ УДАЛЕННЫМИ КОМПЬЮТЕРАМИ В ТЕРМИНАЛЬНОМ РЕЖИМЕ .....	17
3.1. Назначение сервиса Telnet .....	17
3.2. Организация работы сервиса Telnet .....	18
Методические указания .....	19
ГЛАВА 4. СЕРВИС FTP – СИСТЕМА ФАЙЛОВЫХ АРХИВОВ .....	20
4.1. Назначение сервиса FTP .....	20
4.2. Организация работы сервиса FTP .....	21
Методические указания .....	22
ГЛАВА 5. СЕРВИС USENET – СИСТЕМА ТЕЛЕКОНФЕРЕНЦИЙ .....	24
5.1. Назначение сервиса Usenet .....	24
5.2. Организация подписки на конференции .....	25
5.3. Организация работы сервиса Usenet .....	27
Методические указания .....	29
ГЛАВА 6. СЕРВИС E-MAIL – ЭЛЕКТРОННАЯ ПОЧТА .....	30
6.1. Назначение сервиса E-mail .....	30

6.2.	Стандартные функции почтовых клиентов .....	31
6.2.1.	Прием сообщений .....	31
6.2.2.	Локальное хранение сообщений .....	32
6.2.3.	Создание и отправка сообщений .....	33
6.2.4.	Работа с почтовыми вложениями .....	34
6.2.5.	Создание ответных сообщений .....	34
	Методические указания .....	36
ГЛАВА 7. СЕРВИС WWW – ГИПЕРТЕКСТОВАЯ СИСТЕМА ИНТЕГРАЦИИ СЕТЕВЫХ РЕСУРСОВ В ЕДИНОЕ ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО .....		38
7.1.	Назначение сервиса WWW .....	38
7.2.	Средства представления Web-документов .....	40
7.2.1.	Язык HTML .....	40
7.2.2.	Таблицы стилей .....	41
7.2.3.	Сценарии .....	41
7.2.4.	Язык Dynamic HTML .....	42
7.2.5.	Java-апплеты .....	42
7.2.6.	Технология ActiveX .....	43
7.2.7.	Язык XML .....	44
7.3.	Унифицированные указатели ресурса .....	44
7.3.1.	Схемы и механизмы доступа к ресурсу .....	44
7.3.2.	Схема «mailto:» .....	46
7.3.3.	Схема «ftp:» .....	47
7.3.4.	Схемы «http:» и «https:» .....	48
7.4.	Клиентское программное обеспечение .....	49
7.5.	Стандартные функции Web-клиентов .....	49
7.5.1.	Доступ к различным информационным ресурсам .....	49
7.5.2.	Работа с объектами различного типа .....	50
7.5.3.	Обеспечение навигации .....	51
7.5.4.	Кэширование и управление загрузкой объектов .....	51
7.6.	Дополнительные функции Web-клиентов .....	51
7.6.1.	Поддержка маркеров «cookie» .....	51
7.6.2.	Поддержка Java-апплетов и технологии ActiveX .....	52
7.7.	Протоколы взаимодействия клиента и сервера .....	53
	Методические указания .....	54
ГЛАВА 8. СЕРВИС SE – ПОИСКОВЫЕ СИСТЕМЫ .....		57
8.1.	Назначение сервиса SE .....	57
8.2.	Критерии ранжирования документов .....	58

8.3. Организация поиска документов .....	63
8.3.1. Основные поисковые системы .....	63
8.3.2. Поисковая система Яндекс .....	64
8.3.3. Поисковая система Google .....	69
8.3.4. Поисковая система Rambler .....	73
Методические указания .....	77
<b>ГЛАВА 9. СЕРВИС IRC –</b>	
<b>ТЕЛЕКОНФЕРЕНЦИИ РЕАЛЬНОГО ВРЕМЕНИ .....</b>	<b>79</b>
9.1. Назначение сервиса IRC .....	79
9.2. Подключение и настройка mIRC .....	80
9.3. Команды IRC и mIRC .....	84
9.3.1. Список команд IRC и mIRC .....	84
9.3.2. Работа с никами .....	84
9.3.3. Работа с каналами .....	85
9.3.4. Отправка текста .....	87
9.3.5. Управление каналом .....	87
Методические указания .....	88
<b>ГЛАВА 10. СЕРВИС RTVC – ВИДЕОКОНФЕРЕНЦИИ</b>	
<b>РЕАЛЬНОГО ВРЕМЕНИ .....</b>	<b>89</b>
10.1. Назначение сервиса RTVC .....	89
10.2. Организация видеоконференций .....	92
10.2.1. Протоколы семейства H.32x .....	92
10.2.2. Базовая архитектура стандарта H.323 .....	95
10.2.3. Тенденции развития рекомендаций H.323 .....	99
Методические указания .....	101
<b>ЛИТЕРАТУРА .....</b>	<b>103</b>

Учебное издание

КОМАГОРОВ Владимир Петрович

## ТЕХНОЛОГИИ СЕТИ ИНТЕРНЕТ: ПРОТОКОЛЫ И СЕРВИСЫ

Учебное пособие

Научный редактор  
доктор технических наук, профессор *В.А. Силич*  
Редактор *О.М. Васильева*  
Компьютерная верстка *К.С. Чечельницкая*  
Дизайн обложки *О.Ю. Аршинова*

Подписано к печати 27.04.2011. Формат 60x84/16. Бумага «Снегурочка».  
Печать XEROX. Усл. печ. л. 6,22. Уч.-изд. л. 5,63.  
Заказ 596-11. Тираж 35 экз.



Национальный исследовательский Томский политехнический университет  
Система менеджмента качества  
Издательства Томского политехнического университета сертифицирована  
NATIONAL QUALITY ASSURANCE по стандарту BS EN ISO 9001:2008



ИЗДАТЕЛЬСТВО  ТПУ. 634050, г. Томск, пр. Ленина, 30  
Тел./факс: 8(3822)56-35-35, [www.tpu.ru](http://www.tpu.ru)