2. СЛУЖБА WORLD WIDE WEB

[WWW; гипертекст; гиперссылка; HTML; HTTP; URL; CGI]

2.1. Историческая справка

Для решения проблемы устойчивости и надежности каналов связи в 1969 году специалистами из Пентагона была создана крупная децентрализованная компьютерная сеть *Advanced Research Project Agency Network (ARPANET)*. Спустя некоторое время доступ к *ARPANET* получили и гражданские учреждения США, а в 1973 году, с установлением каналов связи с Англией и Норвегией, *ARPANET* преодолела границы США. В 1977 году началось слияние *ARPANET* с другими сетями как внутри США, так и в других развитых странах. В результате такого слияния и родилась глобальная компьютерная сеть *Internet*. В конце 80-х годов Internet, наконец, достиг России.

Internet сделал возможным свободный обмен информацией, невзирая на расстояния и государственные границы. Однако на первом этапе становления его ресурсы были доступны при использовании программного обеспечения, ориентированного лишь на пересылку файлов и неформатированного текста. Физики *Тим Бернерс-Ли* (*Tim Berners-Lee*) и *Робер Кайо* (*Robert Cailliau*) из Женевского *ЦЕРН*а (*CERN*) решили разработать инфраструктуру, позволяющую братьям-физикам по всей Европе обмениваться результатами исследований через Internet в виде привычного для научных работников отформатированного и иллюстрированного текста, включающего ссылки на другие публикации. Так было положено начало *World Wide Web* - Всемирной Информационной Паутине, которая к настоящему времени «*оплела своими сетями практически весь компьютерный мир*» и сделала Internet доступным и привлекательным для миллионов пользователей.

2.2. Структура WWW

Система WWW обеспечивает быстрый и удобный доступ к различным сервисам Internet, например, электронная почта, передача файлов, передача гипертекста, телеконференции и др. Она позволяет организовать удобное взаимодействие через гипертекстовую среду с другими информационными системами, например, с базами данных. Привлекательность WWW заключается также в наличии встроенных в нее возможностей идентификации пользователей и защиты передаваемой информации. Работа WWW базируется на использовании стека протоколов межсетевого обмена TCP/IP (*Transmission Control Protocol*/*Internet Protocol*).

Разнообразная гипертекстовая и гипермедиа информация, расположенная на тысячах Web-серверов во всем мире, является одним из основных видов информационных ресурсов Internet. Разработано множество Web-приложений, предназначенных для практического применения WWW-сервиса в различных областях человеческой деятельности (образование, наука, бизнес, электронная торговля, банковская сфера и т.д.). Наглядным примером перспективного применения WWW-технологии в сфере образования является ее использование в процессе *дистанционного обучения*. Например, на Web-серверах многих вузов и научных организаций высшей школы, наиболее активно использующих возможности WWW-технологии, можно найти ценную и полезную информацию образовательного и познавательного характера, компьютерные учебные пособия, обучающие программы, системы тестирования знаний и другие учебные и методические материалы, которые могут быть использованы в процессе обучения. Представление Web-системы с точки зрения пользователя иллюстрирует схема на рис.2.1.

Все стандарты, имеющие отношение к WWW-технологии, разрабатываются под эгидой консорциума *World Wide Web Consortium* (*W3C*). На его Web-узле (*www.w3.org*, *www.w3c.org.ru*) можно найти спецификации этих стандартов и проекты новых предложений.

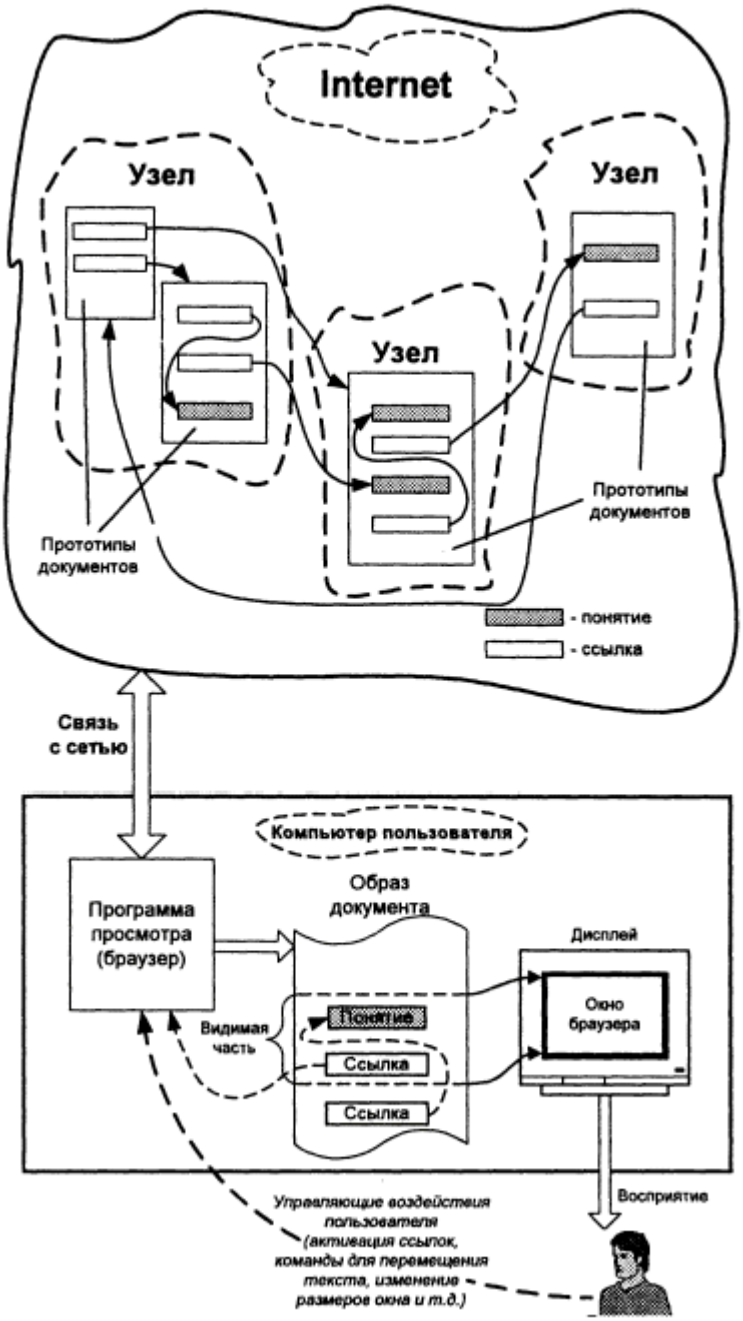

Рис.2.1. Организация доступа пользователя к ресурсам Internet

2.3. Основные компоненты WWW-технологии

WWW представляет собой информационную систему, базирующуюся на использовании понятий «*гипертекст*» и «*гипермедиа*». Гипертекст – это текст со встроенными в него словами (командами) разметки, ссылающимися на другие места этого текста, другие документы, графические изображения и т. д. Выделенные в гипертексте слова называются

гипертекстовыми ссылками (*гиперссылками*). Дальнейшим развитием понятия гипертекста является гипермедиа – синтез гипертекста и мультимедиа. *Гипермедиа-документ* включает не только гипертекст, но и анимацию, аудио- и видеоинформацию.

Структура гипертекстовой системы WWW задается гиперссылками.

Гиперссылка – это адрес другого Web-документа, который тематически, логически или каким-либо иным способом связан с Web-документом, в котором определена ссылка.

Основу WWW-технологии составляют четыре компонента:

- язык разметки гипертекста *HTML* (*HyperText Markup Language*);
- протокол передачи гипертекста *HTTP* (*HyperText Transfer Protocol*);
- универсальный способ адресации ресурсов в сети в виде универсального (унифицированного) указателя ресурса URL (*Uniform Resource Locator*);
- общий шлюзовой интерфейс CGI (*Common Gateway Interface*).

2.3.1. Язык разметки гипертекста HTML

В качестве стандарта для разработки Web-документов принят язык разметки гипертекста HTML – язык компоновки документов и спецификации гиперссылок, используемый для кодирования документов в WWW. Он предназначен для создания форматированного текста, который может включать графические изображения, аудио- и видеоданные, анимацию и гипертекстовые ссылки на другие документы, разбросанные по всему Web-пространству. В основе HTML лежит *теговая модель описания совокупностей элементов*, содержащихся в документе. HTML-документ состоит из текста, который представляет собой содержимое документа, и тегов, определяющих структуру и внешний вид документа при его отображении Web-браузером. Теги – это управляющие маркеры документа, содержащие указания о способе отображения текста и используемые Web-браузером. С помощью тегов создаются ссылки на файлы, содержащие графику, звук, видеоданные, и обозначаются гиперссылки, связывающие одни документы с другими. Гипертекстовая база данных в WWW представляет собой набор текстовых файлов (HTML-документов), написанных на языке HTML и имеющих расширение «*.html*» (или «*.htm*»), например, *data.html*, *index.htm* и т.п.

2.3.2. Протокол передачи гипертекста http

Для обмена данными в WWW используется протокол HTTP (*HyperText Transfer Protocol*). Он обеспечивает передачу гипертекстовой информации с учетом ее специфики. Этот протокол предоставляет пользователю возможность в процессе взаимодействия с сервером получить новый адрес сетевого ресурса, запросить встроенную графику, принять и передать параметры и т.п. Управление в HTTP реализовано с помощью встроенных команд. При работе в Internet для обслуживания HTTP-запросов используется транспортный протокол TCP. Протокол HTTP относится к так называемым «запрос-ориентированным» протоколам. Это означает, что во время сеанса клиент устанавливает соединение и ждет ответа. После отправки ответа сервер инициирует разрыв соединения, поэтому при передаче сложных гипертекстовых страниц соединение может устанавливаться несколько раз.

Говоря о HTTP нельзя не упомянуть протокол *WebDav (Web Distributed Authoring and Versioning*). Этот протокол работает поверх HTTP и служит для доступа к объектам и коллекциям, а также расширяет HTTP набором дополнительных команд. Используется для выполнения основных и расширенных (контроль версий) файловых операций над объектами на серверах, а также для работы с любым типов объектов (не только файлы). Поддерживает одновременную работу над несколькими объектами. Весь перечисленный функционал позволяет в перспективе использовать WebDav в качестве замены FTP.

WebDav сейчас можно встретить при пользовании каким-либо облачным хранилищем. Например, доступ к популярному сервису «Яндекс Диск» можно получить, используя

рассматриваемый протокол. Клиентом могут выступать как специальные приложения, так и стандартные проводники операционных систем (Проводник Windows; *Finder* в *OS X*; *Dolphin* или *Nautilus* в *Linux*).

2.3.3. Универсальный указатель ресурса URL

Универсальный (или унифицированный) указатель ресурса URL (*Uniform Resource Locator*) – адрес в системе WWW, с помощью которого однозначно определяется каждый документ. Он используется для записи гиперссылок и для обеспечения доступа к распределенным ресурсам сети Internet. Синтаксис URL имеет вид:

схема://хост/путь,

где *хост* – имя сервера (хост-машины), с которым необходимо установить соединение; *схема* – протокол, используемый для соединения с сервером (для Web-серверов это протокол http); путь - полное имя документа, затребованного с данного сервера.

Так, например, в конструкции

<A HREF="http://www.riis.ru/FOND_PP/ fond_pp.html">

элемент A, называемый в HTML якорем (*anchor*), использует атрибут HREF, который обозначает гипертекстовую ссылку (*Hypertext REFerence*), для записи этой ссылки в форме URL. Эта ссылка указывает на документ *fond_pp.html* в каталоге FOND_PP на сервере www.riis.ru, доступ к которому осуществляется по протоколу *http*.

В настоящее время широко используется механизм MIME-типов (*Multipurpose Internet Mail Extension*) для идентификации классов внешних ресурсов, поэтому, используя спецификацию URL, в гиперссылке можно указать адрес не только файла HTML, но и файла любого другого формата, для которого есть внешняя программа просмотра. Таким образом, на любой информационный объект можно сослаться из HTML-документа, вызвав его через внешнюю программу просмотра.

Универсальный локатор ресурса используется также для задания адреса какого-либо ресурса в Internet. Имеется несколько схем адресации ресурсов в Internet: HTTP, FTP, FILE, MAILTO, NEWS, TELNET, GOPHER, NNTP, WAIS. Наиболее часто применяются первые три схемы.

Схема HTTP является основной схемой для WWW. В схеме указывается ее идентификатор, адрес машины, TCP-порт (задавать явно необязательно), путь в каталоге сервера, поисковый критерий и метка. На практике обычно используется наиболее простая форма задания адреса с указанием схемы адресации (в это случае – http), доменного имени или IP-адреса машины, имени HTML-файла с указанием полного пути к каталогу, в котором он размещен. Например:

http://www. riis. ru/OI/rosniiis. html или http://194.226.45.45/OI/rosniiis.html

где *http* – протокол обмена информацией в WWW между клиентской машиной и Web-сервером;
www.riis.ru – доменное имя Web-сервера;
194.226.45.45 – IP-адрес Web-сервера;
/OI/rosnii-is.html – полное имя затребованного HTML-файла.

Схема FTP позволяет адресовать файловые архивы FTP из программ-клиентов WWW (Web-браузеров), поддерживающих протокол FTP. В этой схеме возможно указание не только имени схемы и адреса архива, но также идентификатора пользователя и его пароля. Наиболее часто схема FTP используется для доступа к публичным архивам FTP, например:

ftp:// ftp.riis m/pub/index.txt

Этот вариант адреса соответствует идентификатору *anonymous* или *ftp* (анонимный доступ). Идентификатор пользователя и его пароль задаются перед адресом машины:

ftp://gsg:password@ftp.riis.ru/etc

В этом примере имя пользователя – *gsg*, его пароль – *password*, а после знака @ указано имя сервера – *ftp.riis.ru* и каталог – */etc*.

WWW-технология может использоваться как в сетевом, так и в локальном режиме. Для локального режима применяется схема FILE. Следующий пример иллюстрирует обращение к локальному документу (файл *index.html*), размещенному на ПК с операционной системой MS-DOS или MS Windows в каталоге *cop* на диске C:

file:///C:/cop/index.html

2.3.4. Общий шлюзовой интерфейс CGI

Спецификация CGI (*Common Gateway Interface*) была специально разработана для расширения возможностей WWW-технологии за счет подключения всевозможного внешнего программного обеспечения. Основное ее назначение – обеспечение единообразного потока данных между сервером и прикладной программой, которая запускается сервером. CGI – компонент программного обеспечения Web-сервера, который может взаимодействовать с другими программами, работающими на этом сервере. CGI-скрипты – программы, написанные в соответствии со спецификацией CGI на любом языке программирования (*C, C++, C#*, *Pascal* и т. д.) или командном языке (*Shell, Cshell, Perl* и т.д.). CGI – интерфейс, обеспечивающий вызов Web-сервером внешних программ. С помощью CGI Web-сервер может вызвать внешнюю программу и передать в нее пользовательские данные (например, информацию о том, с какой хост-машины пользователь установил соединение, или данные, введенные пользователем в HTML-форму). Эта программа затем обрабатывает полученные данные, а сервер передает результаты ее работы обратно в WWW-браузер. CGI обеспечивает возможность создания сценариев, которые позволяют разрабатывать интерактивные Web-приложения, управляемые пользователем. CGI дает средства динамического создания Web-страниц на основе информации, вводимой пользователем или получаемой из баз данных.

Для расширения возможностей HTML-документов активно используются язык описания сценариев *JavaScript* и объектно-ориентированный язык программирования *Java* (для написания всевозможных программ – от анимации до электронных таблиц, которые можно загружать в локальные машины через WWW).

2.4. Структура программного обеспечения WWW-технологии

Internet-сервис, предоставляемый на основе WWW-технологии, базируется на архитектуре сети «клиент-сервер», поэтому структура программного обеспечения поддержки WWW-сервиса включает две основные части: программу-клиента протокола HTTP (Web-браузер) и программу-сервер протокола HTTP. Они представляют собой минимально необходимый набор программ поддержки WWW-сервиса. Кроме основного программного обеспечения для расширения функциональных возможностей WWW-систем и организации доступа к сетевым ресурсам, отличным от WWW, используются CGI-скрипты, JavaScript-сценарии и Java-*апплеты*. Если серверное и клиентское программное обеспечение (ПО) WWW-систем – это, как правило, стандартное ПО, разрабатываемое известными фирмами, то CGI-скрипты, JavaScript-сценарии и Java-апплеты разрабатываются и пишутся пользователями и администраторами систем.

В WWW функции интерпретатора языка HTML разделены между Web-сервером и Web-клиентом. Сервер, кроме доступа к документам и обработки гипертекстовых ссылок, осуществляет также препроцессорную обработку документов. Программа-клиент WWW выполняет интерпретацию конструкций языка HTML, связанных с представлением информации, и позволяет обмениваться данными с сервером по протоколу HTTP.

Программа-клиент WWW предназначена для того, чтобы соединяться по Internet с удаленными машинами, запрашивать определенные документы, а затем форматировать полученные документы для просмотра на локальной машине.

В настоящее время в качестве Web-клиентов используются браузеры с графическим интерфейсом доступа в WWW, реализованные как для большинства UNIX-систем, так и для систем Wintel. Они обладают прекрасными функциональными возможностями и удобны в работе.

История браузеров с графическим интерфейсом началась с программы *NCSA Mosaic*, которая была создана в Национальном Центре Суперкомпьютерных Приложений Университета штата Иллинойс в Урбана-Шампэйн *(National Center For Supercomputing Applications at the University of Illinois at Urbana-Champaign, NSCA)* в начале 1993 года группой программистов, возглавляемых студентом *Марком Андреессеном* (*Marc Andreessen*). В то время *Mosaic* являлась единственной программой просмотра, написанной на профессиональном уровне, дружественной по отношению к пользователю и работающей на многих типах компьютеров от рабочих станций под управлением операционной системы *UNIX* с графической оболочкой *X Windows* до IBM PC-совместимых компьютеров с *Microsoft Windows* и персональных компьютеров *Apple Macintosh*. Не вызывает сомнений, что именно *Mosaic* открыла в свое время World Wide Web для большинства пользователей[1].

Программа-сервер WWW - это программа, которая принимает запросы от WWW-клиентов и отвечает на них. В качестве ответа может быть возвращен HTML-документ, графическое изображение, аудио- и видеоданные или ответ внешней программы.

В настоящее время Web-серверы существуют для всех типов компьютерных платформ и операционных систем. Наиболее распространенным среди некоммерческих пакетов для UNIX-систем является сервер *Apache* фирмы *Apache Digital*. Хотя этот сервер является некоммерческим, тем не менее, он реализован с учетом спецификаций защиты данных от несанкционированного доступа.

Популярным пакетом, включающим Web-сервер для Wintel-систем, является IIS (Internet Information Server) фирмы Microsoft, работающий в среде ОС Windows xxx Server и поставляемый в комплекте с этой ОС. Основным компонентом IIS является Web-сервер, который позволяет размещать в Internet Web-сайты. IIS поддерживает протоколы HTTP, HTTPS, FTP, POP3, SMTP и др.

ИСТОЧНИКИ ИНФОРМАЦИИ

1. [Электронный ресурс] Сайт W3C (World Wide Web Consortium) – содержит стандарты и их обсуждения: [http://www.w3.org] на 31.12.2013
2. [Электронный ресурс] Сайт DataForce – В этой телеконференции пользователи WWW могут узнать, как установить и использовать программное обеспечение WWW-клиентов [http://www.comp.infosystems.www.users] на 31.12.2013

1 *С особенностями и характеристиками популярных современных Web-браузеров предлагается познакомиться в рамках практического занятия, дополняющего материалы рассматриваемой темы.*