

Лекция1. Основные понятия Информационных технологий.

Интернет-технологии неразрывно связаны с технологией World Wide Web, в переводе "Всемирная паутина", которая получила широкое распространение из-за простоты своих пользовательских интерфейсов. Принцип "жми на то, что интересно", лежащий в основе гипертекста, интуитивно понятен. В технологиях WWW все ключевые понятия просматриваемого документа: слова, картинки - имеют возможность "раскрыться" новым документом, развивающим это понятие. Такой способ представления информации называется "гипертекстом", а документы, представленные в таком виде - "гипертекстовыми документами". Для описания этих документов используется специальный язык - язык описания гипертекстовых документов или HTML (англ. вариант HyperText Markup Language).

1.1. Основные понятия

Использование технологий WWW для обеспечения доступа к каким-либо информационным ресурсам подразумевает существование следующих компонент (см.рис.1-1):

1. TCP/IP - сети с поддержкой базового набора услуг по передаче данных с единой политикой нумерации и маршрутизации, работающим сервисом имен DNS.
2. Выделенного информационного сервера - **WWW-сервера**, обеспечивающего предоставление гипертекстовых документов через IP - сеть в ответ на запросы **WWW - клиентов**.

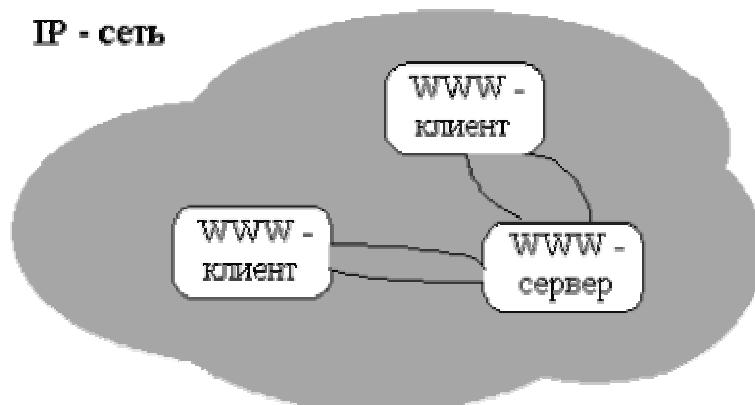


Рис. 1-1

Передаваемые гипертекстовые документы оформляются в стандарте HTML - языке описания гипертекстовых документов. Эти документы могут либо храниться в статическом виде (совокупность файлов на диске), либо динамически компоноваться в зависимости от параметров запроса специальным программным обеспечением

1.2. Назначение TCP/IP - сетей. Общая схема работы.

История и перспективы стека TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP) - это промышленный стандарт стека протоколов, разработанный для глобальных сетей.

Этот протокол был разработан по инициативе Министерства обороны США (Department of Defence, DoD) более 20 лет назад для связи экспериментальной сети ARPAnet с другими сателлитными сетями как набор общих протоколов для разнородной вычислительной среды. Сеть ARPA поддерживала разработчиков и исследователей в военных областях.

Большой вклад в развитие стека TCP/IP внес университет Беркли, реализовав протоколы стека в своей версии ОС UNIX. Широкое распространение ОС UNIX привело и к широкому распространению протокола IP и других протоколов стека. На этом же стеке работает всемирная информационная сеть Internet, чье подразделение Internet Engineering Task Force (IETF) вносит основной вклад в совершенствование стандартов стека.

За долгие годы использования в сетях различных стран и организаций стек TCP/IP накопил большое количество протоколов и сервисов прикладного уровня. К ним относятся такие широко используемые протоколы, как протокол копирования файлов FTP, протокол эмуляции терминала telnet, почтовый протокол SMTP, используемый в электронной почте сети Internet, протокол HTTP, используемый для гипертекстовых сервисов доступа к удаленной информации, таких как WWW. Остановимся несколько подробнее на некоторых из них.

Протокол пересылки файлов **FTP** (File Transfer Protocol) реализует удаленный доступ к файлу. Кроме пересылки файлов протокол FTP предлагает и другие услуги. Так, пользователю предоставляется возможность интерактивной работы с удаленной машиной, например, он может распечатать содержимое ее каталогов. Наконец, FTP выполняет аутентификацию пользователей. Прежде, чем получить доступ к файлу, в соответствии с протоколом пользователи должны сообщить свое имя и пароль. Для доступа к публичным каталогам FTP-архивов Internet парольная аутентификация не требуется, и ее обходят за счет использования для такого доступа предопределенного имени пользователя Anonymous.

Протокол **telnet** обеспечивает передачу потока байтов между процессами, а также между процессом и терминалом. Наиболее часто этот протокол используется для эмуляции терминала удаленного компьютера. При использовании сервиса telnet пользователь фактически управляет удаленным компьютером так же, как и локальный пользователь.

Адресация в IP-сетях

Типы адресов: физический (MAC-адрес), сетевой (IP-адрес) и символьный (DNS-имя)

Каждый компьютер в сети TCP/IP имеет адреса трех уровней:

- Локальный адрес узла, определяемый технологией, с помощью которой построена отдельная сеть, в которую входит данный узел
- IP-адрес, состоящий из 4 байт, например, 109.26.17.100. Он назначается администратором во время конфигурирования компьютеров и маршрутизаторов. IP-адрес состоит из двух частей: номера сети и номера узла. Номер сети может быть выбран администратором произвольно, либо назначен по рекомендации специального подразделения Internet (Network Information Center, NIC), если сеть должна работать как составная часть Internet. Обычно провайдеры услуг Internet получают диапазоны адресов у подразделений NIC, а затем распределяют их между своими абонентами.

Номер узла в протоколе IP назначается независимо от локального адреса узла. Деление IP-адреса на поле номера сети и номера узла - гибкое, и граница между этими полями может устанавливаться весьма произвольно.

- Символьный идентификатор-имя, например, SERV1.IBM.COM. Этот адрес назначается администратором и состоит из нескольких частей, например, имени машины, имени организации, имени домена. Такой адрес, называемый также DNS-именем, используется на прикладном уровне, например, в протоколах FTP или telnet.

Адрес 127.0.0.1 зарезервирован для организации обратной связи при тестировании работы программного обеспечения узла без реальной отправки пакета по сети. Этот адрес имеет название loopback.

Отображение символьных адресов на IP-адреса: служба DNS.

DNS (Domain Name System) - это распределенная база данных, поддерживающая иерархическую систему имен для идентификации узлов в сети Internet. Служба DNS предназначена для автоматического поиска IP-адреса по известному символьному имени узла.

Протокол DNS несимметричен - в нем определены DNS-серверы и DNS-клиенты. DNS-серверы хранят часть распределенной базы данных о соответствии символьных имен и IP-адресов. Эта база данных распределена по административным доменам сети Internet. Клиенты сервера DNS знают IP-адрес сервера DNS своего административного домена и по протоколу IP передают запрос, в котором сообщают известное символьное имя и просят вернуть соответствующий ему IP-адрес.

Если данные о запрошенном соответствии хранятся в базе данного DNS-сервера, то он

сразу посылает ответ клиенту, если же нет - то он посылает запрос DNS-серверу другого домена, который может сам обработать запрос, либо передать его другому DNS-серверу. Клиентские компьютеры могут использовать в своей работе IP-адреса нескольких DNS-серверов, для повышения надежности своей работы.

База данных DNS имеет структуру дерева, называемого доменным пространством имен, в котором каждый домен (узел дерева) имеет имя и может содержать поддомены. Имя домена идентифицирует его положение в этой базе данных по отношению к родительскому домену, причем точки в имени отделяют части, соответствующие узлам домена.

Корень базы данных DNS управляется центром Internet Network Information Center. Домены верхнего уровня назначаются для каждой страны, а также на организационной основе. Для обозначения стран используются трехбуквенные и двухбуквенные аббревиатуры, а для различных типов организаций используются следующие аббревиатуры:

- com - коммерческие организации (например, microsoft.com);
- edu - образовательные (например, mit.edu);
- gov - правительственные организации (например, nsf.gov);
- org - некоммерческие организации (например, fidonet.org);
- net - организации, поддерживающие сети (например, nsf.net).

Каждый домен DNS администрируется отдельной организацией, которая обычно разбивает свой домен на поддомены и передает функции администрирования этих поддоменов другим организациям. Каждый домен имеет уникальное имя, а каждый из поддоменов имеет уникальное имя внутри своего домена. Имя домена может содержать до 63 символов. Каждый хост в сети Internet однозначно определяется своим *полным доменным именем (fully qualified domain name, FQDN)*, которое включает имена всех доменов по направлению от хоста к корню. Пример полного DNS-имени : citint.dol.ru.

1.3. Назначение WWW - сервера. Общая схема работы. Определение

WWW сервер - это такая часть глобальной или внутрикорпоративной сети, которая дает возможность пользователям сети получать доступ к гипертекстовым документам, расположенным на данном сервере. Для взаимодействия с WWW сервером пользователь сети должен использовать специализированное программное обеспечение - браузер (от англ. browser), другое название - программа просмотра.

Схема работы

Рассмотрим более подробно схему работы WWW-сервера. В общем виде она выглядит так:

1. Пользователь сети запускает пакет программного обеспечения, называемый **браузером**, в функции которого входит
 - Установление связи с сервером
 - Получение требуемого документа
 - Отображение полученного документа
 - Реагирование на действия пользователя - доступ к новому документу
 - После запуска браузер по команде пользователя или автоматически устанавливает связь с заданным WWW - сервером и передает ему запрос на получение заданного документа (см рис.3-1).

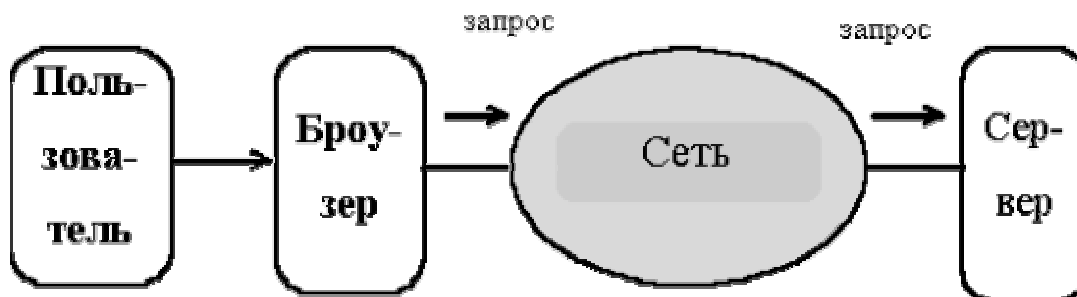


Рис. 3-1

2. WWW сервер ищет запрашиваемый документ и возвращает результаты браузеру (см. рис. 3-2).

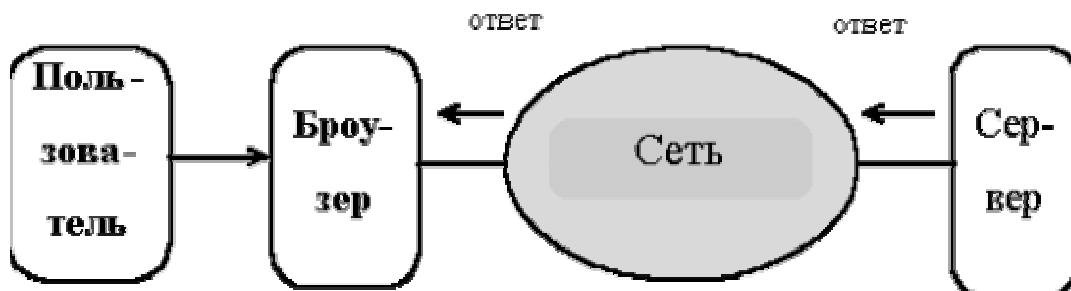


Рис 3-2

3. Броузер, получив документ, отображает его пользователю и ожидает его реакции. Возможные варианты:
 - Ввод адреса нового документа
 - Печать, поиск, другие операции над текущим документом
 - Активизация (нажатие) специальных зон полученного документа, называемых *связями* (link) и ассоциированными с адресом нового документа.

В первом и третьем случае происходит обращение за новым документом.

Адрес

Адрес документа указывается в виде специальной строки, называемой **URL** (Universal Resource Locator). Для протокола HTTP, используемого при взаимодействии WWW клиента и WWW сервера, URL состоит из следующих компонент:

1. Наименование протокола, по которому работает сервер (http).
2. Имя машины - сервера в Internet или ее IP - номер.
3. Порт TCP, обращение к которому обрабатывает сервер.
4. Место (путь) документа на машине - сервере.

Например:

http://www.cnit.nsu.ru:80/welcome.html

Здесь ***http*** означает протокол работы с WWW - сервером

- ***'.'*** - разделитель
- ***"www.cnit.nsu.ru"*** - имя машины - сервера в Internet
- ***"80"*** - номер tcp - порта
- ***/welcome.html*** - путь до документа на машине - сервере

Среда работы сервера

В настоящее время все известные WWW - серверы представляют собой компьютер общего назначения с многозадачной операционной системой. Один или несколько процессов такой системы отвечают за поддержку специфических для WWW - сервера функций. Другие

процессы ОС отвечают за обеспечение других функций, не обязательно связанных с поддержкой технологии WWW (см. рис. 3-3).



Рис. 3-3

Такая структура приводит к тому, что под WWW сервером начинают подразумевать только часть программного обеспечения, единственными функциями которой являются функции WWW сервера, а остальную часть - компьютер, операционную систему, другие процессы, сетевую структуру называют средой работы WWW сервера или платформой. Ниже приведена таблица 3-1, содержащая список наиболее распространенных платформ для WWW - сервера.

Таблица 3-1

Компьютер	Операционная Система
IBM PC	<ul style="list-style-type: none"> • Unix (UnixWare, Open Server, Solaris, BSD, Linux и т.д.) • Microsoft Windows NT • IBM OS/2 • Novell NetWare
Sun SparcStation и SparcServer	<ul style="list-style-type: none"> • SunOS • Solaris
Silicon Graphics серверы и рабочие станции	IRIS

Непосредственные функции сервера.

В простейшем случае гипертекстовый документ представляет собой совокупность файлов. Представление этих файлов как единого документа производится браузером. По каждому файлу документа браузер делает запрос к WWW - серверу. Таким образом, сервер не имеет представления о структуре и составе документов, он отвечает только за выдачу локальных файлов по запросам.

На различных платформах, в различных операционных системах пути файлов выглядят по-разному.

Например:

D:\DOCUMENTS\HTML\INDEX.HTM - в Windows,

/u/data/www/html/index.html - в Unix - системах,

USR:WWW/HTML - в NetWare и т.д.

Путь файла, указываемый в URL, имеет стандартный вид:

/<имя_каталога>/ ... /<имя_каталога>/<имя_файла>

Таким образом, в функции WWW - сервера входит преобразование адреса из внешнего единого формата в платформенно ориентированный внутренний формат. Появляется ряд понятий, специфичных для такого преобразования, необходимых для него.

1. Исходный каталог документов

Это каталог реальной файловой системы сервера, от которого идет вычисление пути, указанного в URL.

Например, если исходным каталогом документов является **D:\Documents\HTML**, то на запрос к этому серверу документа по URL

http://<имя_сервера>/index.htm

будет возвращен файл

D:\Documents\HTML\index.htm

2. Индексный файл.

Для каждого сервера определено имя так называемого индексного файла. Обычно этот файл содержит ссылки на другие файлы данного каталога. Содержимое индексного файла выдается сервером в случае, если в URL указан каталог без конкретного файла.

3. Пользовательский раздел.

Для многопользовательских операционных систем (таких как Unix) ПО WWW - сервера позволяет каждому пользователю предоставлять доступ к своему собственному набору гипертекстовых документов вне основной иерархии (*Исходного каталога документов*, *Синонимов* и т.д.). Этот набор документов должен находиться в собственном (т.н. "домашнем") каталоге пользователя. Для доступа к таким документам в URL перед путем ставится знак тильда и имя пользователя: ~<имя_пользователя>.

Например:

На сервере **Indy.cnit.nsu.ru** создан пользователь с именем fancy и "домашним" каталогом **/home/fancy**. Собственные гипертекстовые документы он хранит в каталоге **/home/fancy/public_html**. При обращении по URL **http://Indy.cnit.nsu.ru/~fancy/start.html**, WWW - сервер будет искать документ **start.html** в каталоге **/home/fancy/public_html**.

Протокол MIME

Протокол MIME (Multipurpose Internet Mail Extensions) - многоцелевое расширение электронной почты, был создан как способ передачи нетекстовой информации: изображений, звука, видео в письмах электронной почты. Механизм оказался удачным, и его перенесли и в on-line сервисы, в том числе WWW. Здесь MIME используется для передачи документов от сервера к клиенту.

В общем виде MIME основывается на передаче вместе с основными данными дополнительной информации, описывающей что это и в каком виде передается. Эта дополнительная информация называется **заголовок MIME**. Базовой частью заголовка является строка, описывающая тип передаваемого сообщения. Формат строки:

Content-Type: <тип_MIME>

Перечень типов MIME (т.е. видов передаваемых данных) постоянно пополняется и может быть дополнен даже пользователем для описания своего собственного вида данных. Формат типа MIME:

<Тип> / <Подтип> [; <параметры>]

Где <Тип> - определяет общий тип данных:

Audio - для звуковых данных

Application - данные, являющиеся входными для какого-либо приложения (программы)

Image - для графических образов

Message - для сообщения, которое само по себе является MIME - документом

Multipart - для сообщения, состоящего из нескольких MIME - документов

Text - для текстовых данных в различном виде

Video - для видеоданных.

<Подтип> - указывает на специфический формат данных типа <Тип>

Например:

text/html - текстовые данные в формате HTML

image/gif - графические данные в формате gifF

<Параметры> - список параметров, необходимых для интерпретации данных.

Протокол HTTP

Протокол HTTP (HyperText Transfer Protocol) определяет язык запросов от WWW - клиента к WWW - серверу. Сам запрос состоит из следующих компонент:

<Заголовок>

<Метод> <Источник / Данные>

где

Заголовок - определяет версию протокола HTTP и другие служебные параметры;

Метод - одно из ключевых слов:

GET - для передачи запросов на выдачу документов

PUT, POST - для передачи данных от клиента к серверу (например, из форм)

Пример запроса:

HTTP/1.1

GET /index.html

Описывает запрос на получение файла index.html из корневого каталога документов сервера.

Интерфейс CGI

Помимо доступа к статическим документам сервера существует возможность получения документов как результата выполнения прикладной программы. Такая возможность реализуется на сервере WWW благодаря использованию интерфейса CGI (Common Gateway Interface). Спецификация CGI описывает формат и правила обмена данными между ПО WWW сервера и запускаемой программой.

Для инициирования CGI необходимо, чтобы в запрашиваемом URL был указан путь до запускаемой программы. ПО WWW сервера исполняет эту программу, передает ей входные параметры и возвращает результаты ее работы, как результат обработки запроса, клиенту. CGI - программой может являться любая программа локальной операционной системы сервера - в двоичном виде или в виде программы для интерпретатора (Basic, SH, Perl и т.д.).

С целью облегчения администрирования CGI - программ, а также для удовлетворения требованиям безопасности CGI - программы группируются в одном или нескольких явно указанных серверу каталогах. По умолчанию это каталог ***cgi-bin*** в иерархии серверных каталогов, однако, его имя и положение могут отличаться.

Например:

клиент, обращающийся к CGI - программе test-query, будет использовать URL ***http://<имя_сервера>/cgi-bin/test-query***

Интерфейс CGI позволяет расширить границы применения WWW - технологии. CGI - программа может обрабатывать сигналы с датчиков установок, взаимодействовать с мощным сервером баз данных, переводить и т.п. Полное описание интерфейса и требований к приложениям, использующих его, приведены в главе 4 настоящего отчета.