

Информационные технологии WWW

План

- Компоненты WWW (HTML, URL, HTTP, CGI)
- Архитектура WWW-технологии
- Методы доступа
- Схема управления ресурсами сервером HTTP

Основные компоненты WWW

- язык гипертекстовой разметки документов HTML;
- универсальный способ адресации ресурсов в сети URL;
- протокол обмена гипертекстовой информацией HTTP;
- универсальный интерфейс шлюзов CGI.

Компонент WWW - HTML

HTML - специальное средство отображения гипертекстовой информации и навигации в гипертекстовой сети.

Пришел на смену стандарту языка разметки - Standard Generalised Markup Language (SGML).

Обычно гипертекстовые системы имеют специальные программные средства построения гипертекстовых связей. Сами гипертекстовые ссылки хранятся в специальных форматах или даже составляют специальные файлы. Такой подход хорош для локальной системы, но не для распределенной на множестве различных компьютерных платформ. В HTML гипертекстовые ссылки встроены в тело документа и хранятся как его часть. Часто в системах применяют специальные форматы хранения данных для повышения эффективности доступа. В WWW документы - это обычные ASCII-файлы, которые можно подготовить в любом текстовом редакторе.

Компонент WWW - URL

Для организации баз данных в WWW требуется Universal Resource Locator (URL).

URL используется в гипертекстовых ссылках и обеспечивает доступ к распределенным ресурсам сети. В URL можно адресовать как другие гипертекстовые документы формата HTML, так и ресурсы e-mail, telnet, ftp, Gopher, WAIS. Различные интерфейсные программы по разному осуществляют доступ к этим ресурсам. Одни, как например Netscape, сами способны поддерживать взаимодействие по протоколам, отличным от протокола HTTP, базового для WWW, другие, как например Chimera, вызывают для этой цели внешние программы. Однако, даже в первом случае, базовой формой представления отображаемой информации является HTML, а ссылки на другие ресурсы имеют форму URL.

Компонент WWW - HTTP

HyperText Transfer Protocol. Данный протокол предназначен для обмена гипертекстовыми документами и учитывает специфику такого обмена. Так в процессе взаимодействия, клиент может получить новый адрес ресурса на сети (relocation), запросить встроенную графику, принять и передать параметры и т. п.

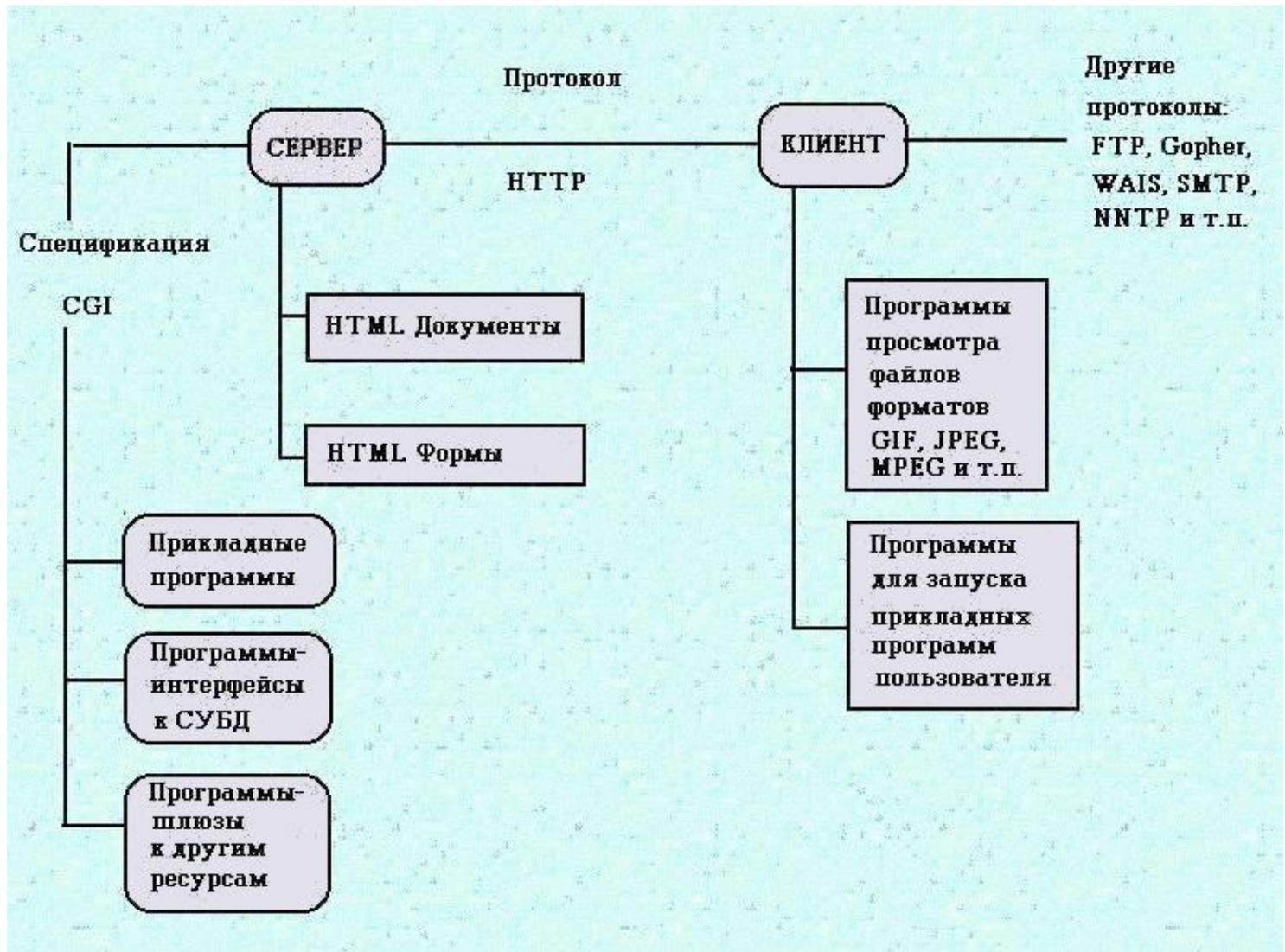
Управление в HTTP реализовано в виде ASCII-команд. Реально разработчик гипертекстовой базы данных сталкивается с элементами протокола только при использовании внешних расчетных программ или при доступе к внешним относительно WWW информационным ресурсам, например базам данных.

Компонент WWW - CGI

CGI (Common Gateway Interface) была специально разработана для расширения возможностей WWW за счет подключения всевозможного внешнего программного обеспечения. Такой подход логично продолжал принцип публичности и простоты разработки и наращивания возможностей WWW.

Предложенный и описанный в CGI способ подключения не требовал дополнительных библиотек и буквально ошеломлял своей простотой. Сервер взаимодействовал с программами через стандартные потоки ввода/вывода, что упрощает программирование до предела. При реализации CGI чрезвычайно важное место заняли методы доступа, описанные в HTTP. И хотя реально используются только два из них (GET и POST), опыт развития HTML показывает, что сообщество WWW ждет развития и CGI по мере усложнения задач, в которых будет использоваться WWW-технология.

Архитектура WWW - технологии



WWW построена по схеме "клиент-сервер". На рисунке представленном выше показано, как разделены функции в этой схеме. Программа-клиент выполняет функции интерфейса пользователя и обеспечивает доступ практически ко всем информационным ресурсам Internet.

Фактически, клиент - это интерпретатор HTML. И как типичный интерпретатор, клиент в зависимости от команд (разметки) выполняет различные функции. В круг этих функций входит не только размещение текста на экране, но обмен информацией с сервером по мере анализа полученного HTML-текста, что наиболее наглядно происходит при отображении встроенных в текст графических образов. При анализе URL-спецификации или по командам сервера клиент запускает дополнительные внешние программы для работы с документами в форматах, отличных от HTML, например GIF, JPEG, MPEG, Postscript и т. п. Вообще говоря для запуска клиентом программ независимо от типа документа была разработана программа Lurcher, но в последнее время гораздо большее распространение получил механизм согласования запускаемых программ через MIME-типы.

Другую часть программного комплекса WWW составляет сервер протокола HTTP, базы данных документов в формате HTML, управляемые сервером, и программное обеспечение, разработанное в стандарте спецификации CGI. В настоящее время наиболее популярны сервер для MS-Windows и Apache-сервер для Unix-платформ.

База данных HTML-документов - это часть файловой системы, которая содержит текстовые файлы в формате HTML и связанные с ними графику и другие ресурсы. Особое внимание хотелось бы обратить на документы, содержащие элементы экранных форм. Эти документы реально обеспечивают доступ к внешнему программному обеспечению.

Прикладное программное обеспечение, работающее с сервером, можно разделить на программы-шлюзы и прочие. Шлюзы - это программы, обеспечивающие взаимодействие сервера с серверами других протоколов, например ftp, или с распределенными на сети серверами Oracle. Прочие программы -- это программы, принимающие данные от сервера и выполняющие какие-либо действия: получение текущей даты, реализацию графических ссылок, доступ к локальным базам данных или просто расчеты.

Методы доступа

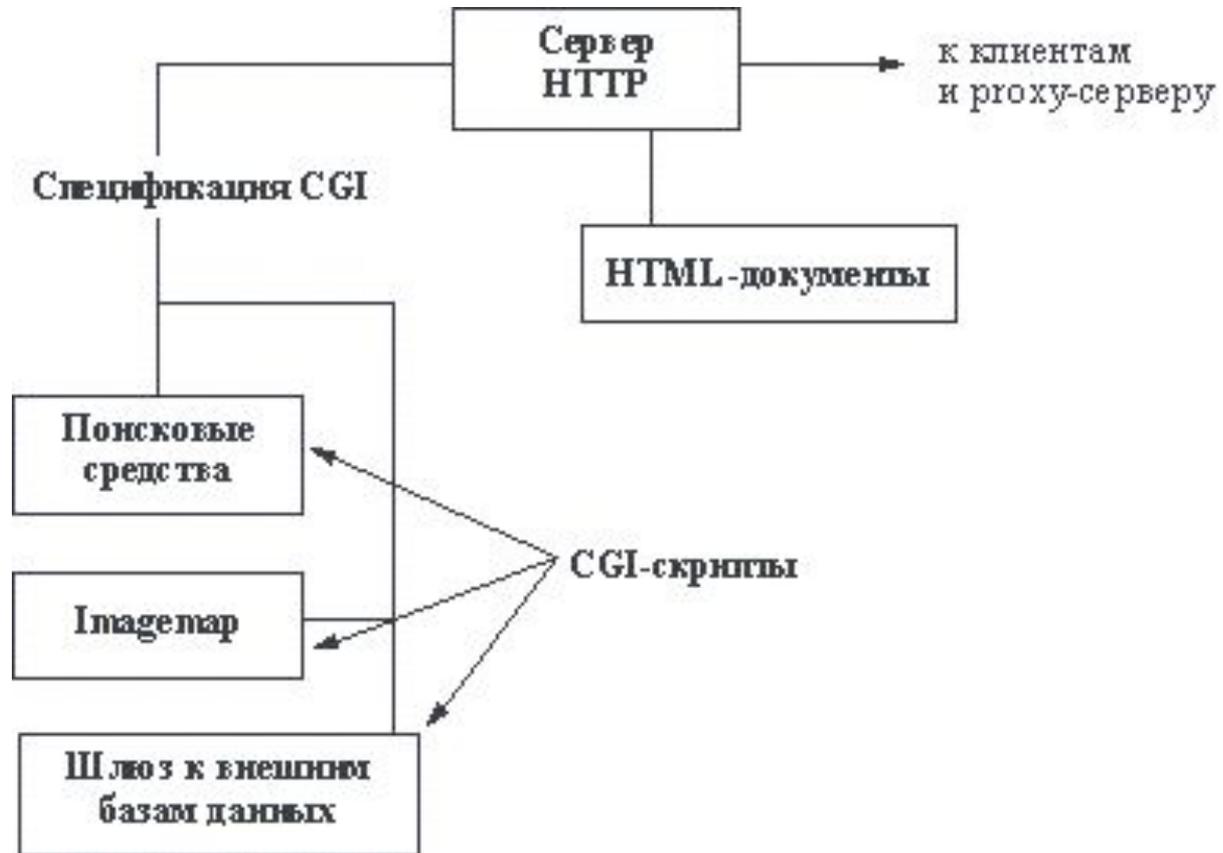
Существует три метода доступа: POST, GET, HEAD.

GET - метод, позволяющий получить данные, заданные в форме URI в запросе ресурса. Если ссылаются на программу, то возвращается результат выполнения этой программы, но не текст программы. Дополнительные данные, которые надо передать для обработки, кодируются в запрос ресурса. Имеется разновидность метода GET - условный GET. При использовании этого метода сервер ответит на запрос только в том случае, если будут выполнены условия передачи. Это позволяет разгрузить сеть, избавив ее от передачи ненужной информации. Условие указывается в поле "if-Modified-Since" заголовка запроса. При использовании метода GET в поле тела ресурса возвращается собственно затребованная информация (текст HTML-документа, например).

HEAD - метод, аналогичный GET, но не возвращает тела ресурса. Используется для получения информации о ресурсе. Условного HEAD не существует. Данный метод используется для тестирования гипертекстовых ссылок.

POST - этот метод разработан для передачи большого объема информации на сервер. Им пользуются для аннотирования существующих ресурсов, отправки почтовых сообщений, работы с формами интерфейсов к внешним базам данных и внешним исполняемым программам. В отличие от GET и HEAD, в POST передается тело ресурса, которое и является информацией из поля форм или других источников ввода. В первых версиях протокола были определены и другие методы доступа (DELETE, например), но они не нашли должного применения. Многие функции, которые возлагали на эти методы, можно успешно выполнять через POST.

Схема управления ресурсами сервером HTTP



Контрольные вопросы

1. Перечислите и опишите компоненты WWW.
 2. Дайте описание архитектуре WWW-технологии.
 3. Перечислите и опишите методы доступа.
 4. Поясните схему управления ресурсом сервером HTTP
- .