

## 7.5. Применение гипертекстовых технологий в глобальных сетях

Областью самого массового применения гипертекстовых технологий является сетевая служба World Wide Web (WWW - всемирная паутина) глобальной сети Internet. Служба WWW предоставляет набор услуг Internet, позволяющий просмотреть любые данные, хранящиеся на компьютерах этой сети через систему связывающих их гиперссылок.

Можно выделить четыре составляющих элемента сетевой службы WWW,

**Язык гипертекстовой разметки HTML** (HyperText Markup Language) является стандартным языком, предназначенным для создания гипертекстовых документов в среде WWW. Он был предложен Тимом Бернерсом-Ли в 1989 г. в качестве одного из компонентов технологии разработки распределенной гипертекстовой системы World Wide Web.

Основная идея языка HTML заключается в следующем: текст документа дополняется информацией, задающей макетную и логическую структуры документа, таким образом, структура HTML-документа включает два функциональных элемента:

<b>Собственно текст</b> , т.е. данные, составляющие содержимое документа	<b>Тэги (markup tags)</b> , называемые также флагами разметки, - специальные конструкции языка HTML, используемые для разметки документа и управляющие его отображением
--	---

Тэги языка HTML управляют форматированием текста и определяют, в каком виде будет представлен текст, какие его компоненты будут исполнять роль гипертекстовых ссылок, какие графические или мультимедийные объекты должны быть включены в документ. Графическая и звуковая информации, включаемые в HTML-документ, хранятся в отдельных файлах.

По своему значению тэги близки к понятию "скобок". В большинстве случаев они используются парами. Пара состоит из открывающего (start tag) и закрывающего (end tag) тэгов. Идентификаторы тэгов заключаются в треугольные скобки, имя закрывающего тэга отличается от имени открывающего лишь тем, что перед ним ставится наклонная черта:

<имя тэга> - открывающий тэг;

</имя тэга> - закрывающий тэг.

Например, тэги <html> и соответственно </html> открывают и закрывают сам HTML-документ, а текст, стоящий между тэгами <b> и </b>, будет выделен полужирным шрифтом.

Пример фрагмента HTML-документа приведен в [табл. 7.2](#).

Таблица 7.2. Пример разметки HTML-документа

Текст фрагмента	Разметка фрагмента в формате HTML-документа	Описание разметки
Глобальная вычислительная	<p>	<p> — начало абзаца;

<b>сеть Internet</b> предоставляет следующие услуги: <ul style="list-style-type: none"> <li>• электронная почта;</li> <li>• телеконференции;</li> <li>• служба WWW</li> </ul>	<b>&lt;b&gt;</b> Глобальная вычислительная сеть	<b>&lt;b&gt;</b> и <b>&lt;/b&gt;</b> — начало и конец текста, который должен выделяться полужирным шрифтом;	
	<b>Internet&lt;/b&gt;</b> предоставляет следующие услуги: <b>&lt;ul&gt;</b>	<b>&lt;li&gt;</b> электронная почта; <b>&lt;/li&gt;</b>	<b>&lt;ul&gt;</b> — начало нумерованного списка;
	<b>&lt;li&gt;</b> телеконференции; <b>&lt;/li&gt;</b>	<b>&lt;li&gt;</b> и <b>&lt;/li&gt;</b> — начало и конец элемента списка;	
	<b>&lt;li&gt;</b> служба WWW. <b>&lt;/li&gt;</b> <b>&lt;/ul&gt;</b> <b>&lt;/p&gt;</b>	<b>&lt;/ul&gt;</b> — конец списка; <b>&lt;/p&gt;</b> — конец абзаца	

Существует два способа создания гипертекстовых документов:

- **создание документа с помощью HTML-редактора.** Этот способ позволяет создавать документы для WWW без знания языка HTML. HTML-редакторы автоматизируют создание гипертекстовых документов и избавляют разработчика от рутинной работы. Однако их возможности ограничены, они сильно увеличивают размер получаемого файла и не всегда полученный с их помощью результат соответствует ожиданиям.
- **создание и разметка документа при помощи обычного текстового редактора.** При этом способе разработчик вручную вставляется в текст команды языка HTML.

Для просмотра HTML-документов существуют специальные программы, которые называются *браузеры*. Они интерпретируют тэги разметки документа и располагают текст и графику на экране соответствующим образом.

Для записи гипертекстовых ссылок в системе WWW была разработана специальная форма, которая называется **универсальный способ адресации ресурсов в сети URL (Uniform Resource Locator)**. Указатель ресурса URL - это адрес в системе WWW, при помощи которого однозначно определяется каждый документ. Он используется для записи гиперссылок и для обеспечения доступа к распределенным ресурсам сети Internet. Синтаксис URL имеет такой вид:

**схема://хост/путь,**

где:

- **схема** - это протокол, используемый для соединения с данным хост-компьютером;
- **хост** - это имя хост-компьютера, с которым необходимо установить соединение;
- **путь** - полное имя документа, затребованного с данного сервера.

Например, в указателе ресурса <http://www.dlib.org/dlib.html>

**http** - означает название протокола для соединения с компьютером;

**www.dlib.org** - имя компьютера;

**dlib.html** - имя файла на этом компьютере.

Неформально этот URL может интерпретироваться следующим образом: "Используя HTTP-протокол, соединиться с компьютером www.dlib.org и получить файл dlib.html".

В настоящее время активно используются следующие схемы (протоколы) адресации ресурсов в Internet, приведенные в [табл. 7.3](#):

*Протокол передачи гипертекста HTTP* (HyperText Transfer Protocol) используется для обмена данными в системе WWW. 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.

Схема доступа (протокол)	Описание
file	Имя файла в компьютере
ftp	Протокол передачи файлов
gopher	Протокол службы Gopher
http	Протокол передачи гипертекста
mailto	Адрес электронной почты
news	Новости телеконференции USENET
telnet	Сеанс удаленного доступа telnet

Каждый запрос клиента и ответ сервера WWW состоит из трех частей:

- строка запроса (ответа);
- раздел заголовка;
- тело запроса.

Структура запроса клиента и ответа сервера

*Запрос клиента:*

- Строка запроса включает HTTP-команду, называемую методом, адрес документа и номер версии протокола HTTP. Например, строка запроса

**GET /index.html HTTP/1.0**

обозначает, что для запроса используется метод GET, которым с помощью версии 1.0 протокола HTTP запрашивается документ *index.html*.

Метод - это HTTP-команда, с которой начинается первая строка запроса клиента. Метод сообщает серверу о цели запроса. Для HTTP определены три основных метода: GET, HEAD и POST. Характеристика основных методов запроса клиента приведена в [табл. 7.4](#).

Название метода	Характеристика метода
Метод GET	Метод GET запрашивает информацию, расположенную на сервере по указанному адресу URL. GET — наиболее распространенный метод поиска информации с помощью браузеров. Результат запроса GET может представлять собой:

	<ul style="list-style-type: none"> <li>• файл;</li> <li>• результат выполнения программы;</li> <li>• выходную информацию аппаратного устройства и т. д.</li> </ul>
Метод HEAD	<p>Метод HEAD запрашивает только информацию заголовка о файле или ресурсе. Этот метод используется, когда клиент хочет найти информацию о документе, не получая его. Например, клиент может затребовать следующую информацию:</p> <ul style="list-style-type: none"> <li>• время изменения документа;</li> <li>• размер документа;</li> <li>• тип документа;</li> <li>• тип сервера и т. д.</li> </ul>
Метод POST	<p>Метод POST позволяет посылать на сервер данные в запросе клиента. Эти данные направляются в программу обработки данных, к которой сервер имеет доступ. Метод POST может использоваться во многих приложениях. Например, его можно применять для передачи входных данных для:</p> <ul style="list-style-type: none"> <li>• сетевых служб (таких как телеконференции);</li> <li>• программ с интерфейсом в виде командной строки;</li> <li>• аннотирования документов на сервере;</li> <li>• выполнения операций в базах данных</li> </ul>

**Раздел заголовка** запроса является необязательным и может содержать информацию о конфигурации клиента и данные о форматах документов, которые он может принимать. Завершается раздел заголовка пустой строкой.

**Тело запроса** содержит дополнительные данные, которые используются программами обработки на WWW-сервере. Тело запроса также не является обязательным.

*Ответ сервера.*

**Строка ответа** включает версию протокола HTTP, которой данный сервер пользуется для передачи ответа, код состояния и описание.

Код состояния - это трехразрядное число, обозначающее результат обработки сервером запроса клиента. Описание, следующее за кодом состояния, представляет собой просто понятный для человека текст, поясняющий код состояния. Например, строка состояния

### **HTTP/1.0 200 ОК**

обозначает, что сервер для ответа использует версию *HTTP 1.0*. Код состояния *200* означает, что запрос клиента был успешным и затребованные данные будут переданы после заголовков.

*Раздел заголовка ответа* содержит данные о самом сервере и затребованном документе. Завершает заголовок пустая строка.

Если запрос клиента успешен, то в теле ответа посылаются затребованные данные. Если запрос клиента удовлетворить нельзя, то передаются дополнительные данные в виде понятного для пользователя разъяснения причин, по которым сервер не смог выполнить данный запрос.

**Универсальный интерфейс шлюзов CGI (Common Gateway Interface)** был специально разработан для расширения возможностей WWW-технологии за счет подключения всевозможного внешнего программного обеспечения. Основное ее назначение - это обеспечение единообразного потока данных между сервером и прикладной программой, которая запускается сервером. CGI - это компонент программного обеспечения Web-сервера, который может взаимодействовать с другими программами, работающими на этом сервере. 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.