Лабораторная работа №1

Язык разметки гипертекста HTML. Начало работы**.**

**ЦЕЛЬ РАБОТЫ: Составить общее представление о гипертекстовой форме хранения информации. Получить базовые навыки по созданию и интерпретации информационных ресурсов с помощью языка разметки гипертекста HTML. Установка и запуск Веб-сервера.**

1. Понятие гипертекста

В 1989 году, когда Тим Бернерс-Ли представил систему WWW, в мире информационных технологий наблюдался повышенный интерес к новому и модному в то время направлению - гипертекстовым системам. Сама идея, но не термин, была введена В.Бушем (Vannevar Bush) в 1945 году в предложениях по созданию электромеханической информационной системы Memex. Несмотря на то, что Буш был советником по науке президента Рузвельта, идея не была реализована. В 1965 году Т.Нельсон (Ted Nelson) ввел в обращение сам термин «*гипертекст»*, развил и даже реализовал некоторые идеи по работе с «нелинейными» текстами.

Идея гипертекстовой информационной системы состоит в том, что пользователь имеет возможность просматривать документы (страницы текста) в произвольном порядке, а не последовательно, как это принято при чтении книг. Поэтому Т.Нельсон и определил «гипертекст» как *«нелинейный текст»*. Достигается это путем создания специального механизма связи различных страниц текста при помощи гипертекстовых ссылок, т.е. у обычного текста есть ссылки типа "следующий-предыдущий", а у гипертекста можно построить еще сколь угодно много других ссылок. Любимыми примерами специалистов по гипертексту являются энциклопедии, Библия, системы типа "Help". Дальнейшее развитие этого подхода приводит к расширению понятия гипертекста за счет других информационных ресурсов, включая графику, аудио- и видеоинформацию, до понятия *«гипермедиа»*.

2. HTML – общие положения

Язык гипертекстовой разметки HTML (HyperText Markup Language) был предложен Тимом Бернерсом-Ли в 1989 году в качестве одного из компонентов технологии распределенной гипертекстовой системы World Wide Web.

Разработчики HTML пытались решить две задачи:

* дать дизайнерам гипертекстовых баз данных простое средство создания документов;
* сделать это средство достаточно мощным, чтобы отразить имевшиеся на тот момент представления об интерфейсе пользователя гипертекстовых баз данных.

Первая задача была решена за счет выбора тэговой модели описания документа. Такая модель широко применяется в системах подготовки документов для печати. Примером такой системы является хорошо известный язык разметки научных документов TeX, предложенный Американским Математическим Обществом. К моменту создания HTML существовал стандарт языка разметки печатных документов - SGML (Standard Generalized Markup Language), который и был взят в качестве основы HTML.

Вторым важным моментом, повлиявшим на судьбу HTML, стал выбор в качестве элемента гипертекстовой базы данных обычного текстового файла, который хранится средствами файловой системы операционной среды компьютера. Такой выбор был сделан под влиянием следующих факторов:

* такой файл можно создать в любом текстовом редакторе на любой аппаратной платформе в среде любой операционной системы.
* к моменту разработки HTML существовал американский стандарт для разработки сетевых информационных систем - Z39.50, в котором в качестве единицы хранения указывался простой текстовый файл в кодировке LATIN1, что соответствует US ASCII.

Таким образом, гипертекстовая база данных в концепции WWW - это набор текстовых файлов, написанных на языке HTML, который определяет форму представления информации (разметка) и структуру связей этих файлов (гипертекстовые ссылки).

Такой подход предполагает наличие еще одной компоненты технологии - интерпретатора языка. В World Wide Web функции интерпретатора разделены между сервером гипертекстовой базы данных и интерфейсом пользователя.

Сервер, кроме доступа к документам и обработки гипертекстовых ссылок, осуществляет также препроцессорную обработку документов, в то время как интерфейс пользователя осуществляет интерпретацию конструкций языка, связанных с представлением информации. В качестве интерфейса пользователя обычно выступает средство просмотра HTML-документов – Web-браузер.

Первая версия языка (HTML 1.0) была направлена на представление языка как такового – описание его возможностей носило скорее рекомендательный характер. Вторая версия языка (HTML 2.0) фиксировала практику использования конструкций языка. В версии ++ (HTML++) был расширен набор элементов, реализующих отображение научной информации и таблиц, кроме того улучшился стиль компоновки изображений и текста. Версия 3.0 была призвана упорядочить все нововведения и согласовать их с существующей практикой. Версия 3.0 так и не была закончена – язык развивался настолько быстро, что она стала неактуальной еще до официального завершения работ. «Масла в огонь подлили» фирмы, разрабатывающие WEB-браузеры, – каждая из них старалась включить в HTML целый ряд собственных расширений, призванных «существенно упростить» труд создателей HTML-ресурсов. Но все это вызвало лишь дополнительную путаницу. Компромиссным решением стал выпуск версии 3.2 (до сих пор являющейся самой стабильной и универсальной), ибо в ней были учтены как «нововведения» сторонних разработчиков, так и пожелания комитета стандартизации W3C (организации, занимающейся поддержкой и развитием системы WWW). На сегодняшний день последней официальной версией HTML является 4.0, носящая негласное название Dynamic HTML.

3. Принципы построения и интерпретации HTML

Для разметки документа с помощью HTML используется специальная тэговая модель. *Тэговая модель* описывает документ как совокупность элементов, каждый из которых окружен *тэгами*, первый из которых является открывающим, а второй соответственно закрывающим (в некоторых случаях он может отсутствовать). По своему значению тэги близки к понятию операторных скобок "begin/end" в универсальных языках программирования. Тэгами обозначается область действия правил интерпретации конкретного элемента документа. В качестве типичного примера можно привести тэги элемента Italic, определяющего область отображения текста курсивом.

Текст на языке HTML:

**Текст, следующий за словом "Italic" <I>отображается курсивом</I>.**

Текст, отображаемый программой интерпретации:

**Текст, следующий за словом "Italic" *отображается курсивом*.**

В приведенном выше примере элемент текста, который должен быть выделен курсивом, заключен между тэгом начала стиля "Italic" - <I> и тэгом конца стиля - </I>. Общая схема построения элемента документа в формате HTML может быть записана в следующем виде:

**"элемент" := <"имя элемента" "список атрибутов"> содержание элемента </"имя элемента">**

Конструкция перед содержанием элемента называется тэгом начала элемента, а конструкция, расположенная после содержания элемента, – тэгом конца элемента. Тэг может обладать набором атрибутов, записываемых внутри угловых скобок в виде ***атрибут=значение***, где «значение» представляет собой строку, заключенную в кавычки (хотя на практике это и не всегда обязательно). Запись тэгов и атрибутов является «регистронечувствительной», хотя правилом хорошего тона является «выделение» наиболее значимых элементов (например, элементов базовой разметки страницы – HTML, HEAD, BODY и т.д.) символами верхнего регистра.

Пример тэга с тремя атрибутами:

**<table border="0" cellpadding="0" cellspacing="0">**

Символы, не входящие ни в один из элементов, воспринимаются как «текстовый поток» и не интерпретируются. Специальные символы (например – символ перевода строки) не имеют в HTML никакого значения. По умолчанию, перевод строки происходит при достижении границы экрана. Для управления этим процессом служат специальные элементы языка.

Структура гипертекстовой сети задается гипертекстовыми ссылками. Гипертекстовая ссылка - это адрес другого HTML документа, который тематически, логически или каким-либо другим способом связан с документом, в котором ссылка определена.

Для записи гипертекстовых ссылок в системе WWW используется форма URI. Вот типичный пример использования этой записи:

**Этот текст содержит**

**<A HREF="http://polyn.net.kiae.su/altai/index.html">**

**гипертекстовую ссылку</A>.**

В приведенном выше примере элемент "A" (который в HTML называют якорем (anchor) – он «цепляется» за другие документы) использует атрибут "HREF" (**H**ypertext **REF**erence) для записи гипертекстовой ссылки в форме URI. Данная ссылка указывает на документ с именем "index.html" в директории "altai" на сервере "polyn.net.kiae.su", доступ к которому осуществляется по протоколу "http". Кроме того, с помощью названного элемента можно устанавливать метки в теле документа. Для этого служит атрибут name, задающий имя метки, на которую можно ссылаться в URI (имя метки завершает URI-адрес и отделено от основной его части символом “#”)

Структура HTML-документа позволяет использовать вложенные друг в друга элементы. Собственно, сам документ - это один большой элемент с именем "HTML":

**<HTML> Содержание документа </HTML>**

Сам элемент «HTML», или гипертекстовый документ, состоит из двух частей: заголовка документа (HEAD) и тела документа (BODY):

**<HTML>**

**<HEAD>**

**Содержание заголовка**

**</HEAD>**

**<BODY>**

**Содержание тела документа**

**</BODY>**

**</HTML>**

Заголовок документа содержит данные, большая часть которых не демонстрируется пользователю – т.е. является служебной. Обычно это информация об авторских правах, инструментарии, типах представления данных и т.д.

Рассмотрим пример классического документа:

**<HTML>**

**<!--**

**Author: Pavel Khramtsov**

**Date: January 21, 1996**

**-->**

**<HEAD>**

**<TITLE>This is a Banner</TITLE>**

**</HEAD>**

**<BODY BACKGROUND=”www\_wall.jpg” VLINK=”#0000FF” LINK=”#FF0000”>**

**<CENTER>**

**<TABLE>**

**<TR><TD><IMG SRC="interne0.jpg"></TD>**

**<TD ALIGN=”CENTER”>**

**<H3>Администрирование Internet</H3>**

**<I>Центр Информационных Технологий, 1996.</I>**

**</TD></TR>**

**</TABLE>**

**</CENTER>**

**</BODY>**

**</HTML>**

Все, что расположено между <HTML> и </HTML> - это документ. Содержимое элемента HEAD определяет заголовок документа (в примере он состоит из одного элемента TITLE). Каждый документ в системе WWW имеет свое название, которое указывается в элементе TITLE заголовка документа. Им обычно озаглавливается рабочее окно интерфейса (браузера).

Контейнер BODY содержит тело документа (иными словами, сам документ). В качестве фона в этом элементе определено графическое изображение из файла www\_wall.jpg. Имя файла - "www\_wall.jpg" - задано частной формой URI, которая не определяет полного адреса ресурса в сети. Затем в документе определена таблица, состоящая из двух ячеек. В одной ячейке находится графическое изображение interne0.jpg, в то время как в другой - текстовый фрагмент. Текст определен как заголовок третьего уровня, который должен отображаться стилем Italic.

Информация об авторе приведенного документа и дате создания заключена в пару тэгов, называемых комментариями. Подобные элементы HTML-страницы не демонстрируются средствами просмотра (браузерами).

3. Элементы языка

Прежде всего, элементы HTML делятся на *логические* и *физические*. Логические – появились вместе со спецификацией языка и представляют собой конструкции, визуальное представление которых полностью зависит от браузера. Т.е. один и тот же элемент в разных средствах просмотра HTML-ресурсов может отображаться совершенно по-разному. Основная задача логических элементов – описать структурную единицу разметки и ее назначение (но не внешний вид!).

Примеры логических элементов:

**<ЕМ>...</ЕМ>** Выделенный текст

**<Hx>...</Hx>** Заголовок уровня x.

**<STRONG>...</STRONG>** Особо выделенный текст

**<CITE>...</CITE>** Текст ссылки на книгу или документ (цитата)

**<CODE>...</CODE>** Фрагмент исходного кода

**<DFN>...</DFN>** Определенное оформление слова

**<KBD>...</KBD>** Ввод с клавиатуры

**<SAMP>...</SAMP>** Пример текста

**<VAR>...</VAR>** Имя переменной

Физические элементы HTML явно описывают внешний вид единицы разметки. Примером их использования является указание кегля или начертания шрифта текста, а также задание точных размеров любого из элементов документа.

Примеры физических элементов:

**<B>...</B>** Полужирный

**<U>...</U>** Подчеркнутый

**<I>...</I>** Курсив

**<ТТ>...</ТТ>** Моноширинный

**<S>...</S>** Зачеркнутый

**<SUB>...</SUB>** Нижний индекс

**<SUP>...</SUP>** Верхний индекс

**<FONT>...</FONT>** Шрифт

Использование логических элементов является предпочтительным, так как это делает формируемый документ платформеннонезависимым (ответственность за правильность разметки перекладывается на браузер).

Ход лабораторной работы (практическая часть)

1. Ознакомиться с материалами методического указания.\*
2. Составить HTML-документы (2-3) произвольного содержания (на выбранную тему), содержащий элементы разметки следующих функциональных групп: гиперссылки, графические изображения, таблицы, заголовки, текстовые блоки (абзацы), списки (что-нибудь на свой выбор). В документе должны присутствовать как логические, так и физические элементы языка.
3. Ознакомиться с инструкцией по установке и запуску веб-сервера (в аудиториях лаборатории уже установлен Denwer (Apache, php, mysql), в своей работе можете использовать удобные для вас веб-сервера и сборки: nginx, LAMP, XAMPP). Создать свой веб-ресурс (залить туда док-ты созданные в п.2) на сервере и проверить работу.
4. Проверить работу интерпретатора php:

- расширение страницы поменять на .php и вставить внутри теги <?php ?>. Внутри тегов написать простой вывод текста инструкцией echo, например:

echo 'Hello world';

- проверить вывод текста в html-документ, проверить вывод тегов (например таблиц, ссылок и т.д.).

\* - При наличии достаточных (на ваш взгляд) знаний в предметной области, чтение можно пропустить и перейти непосредственно к практической части.