ГИПЕРТЕКСТОВЫЕ СПОСОБЫ ХРАНЕНИЯ И ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ

6.1. Понятие и основные элементы гипертекстовой технологии

Гипертекстовая технология — это технология преобразования текста из линейной формы в иерархическую, поэтому использование гипертекстовой технологии (по сравнению с представлением информации в обычной книге) позволяет кардинально изменить способ просмотра и восприятия информации. Так, читая текст в книге, мы просматриваем его последовательно, страница за страницей. И если в процессе чтения мы встретим термин, значение которого объяснялось раньше, то в этом случае нам придется листать страницы книги в обратном порядке до тех пор, пока не найдем нужное нам определение непонятного термина. Использование же гипертекстовой технологии позволяет значительно упростить работу с текстом и найти нужное определение за считанные секунды [11].

В настоящее время гипертекстовая технология широко используется для построения подсистем помощи пользователям при работе с диалоговыми компьютерными программами, а также для построения различных справочников, энциклопедий.

С развитием компьютерных средств мультимедиа гипертекст начал превращаться в более наглядную информационную форму, получившую название гипермедиа — эта информационная форма содержит не только текст, но и графику, видеоинформацию и звуки.
В простейшем случае технология построения гипертекста включает следующие пять основных этапов:

1) разделение текста на отдельные главы или темы;
2) выбор основного маршрута чтения гипертекста и расстановка ссылок, ведущих пользователя по темам последовательно, в соответствии с основным маршрутом;
3) определение дополнительных маршрутов чтения, которые могут оказаться удобными читателю, и расстановка ссылок, позволяющих осуществить логичный переход от основного маршрута к дополнительным;
4) выявление и написание недостающих частей текста, которые требуются для логичного следования по маршрутам чтения;
5) связь ссылок с существующими темами.

Гипертекстовая форма представления информации позволяет не только сделать текст структурированным, но и организовать моментальный переход читателя к интересующим его разделам с помощью ссылок (рис. 6.1). В результате с помощью гипертекста читателю предоставляется возможность самостоятельно выбирать порядок работы с материалом, изменять маршрут непосредственно в процессе чтения.

Простота концепции гипертекста обусловливает и формальную простоту общей принципиальной технологии создания гипертекстов. Имея простейшую систему построения гипертекстов, можно быстро собрать из нескольких текстовых фрагментов гипертекст и формально получить самостоятельную гипертекстовую информационную систему, программный продукт или подсистему подсказки. Однако в силу видимой простоты гипертекстовой технологии легко создать гипертекстовую информационную систему с низким качеством.

Гипертексты обладают определенной семантической (смысловой) сетевой структурой. При многократном просмотре, если гипертекст используется как учебник, эта структура будет сильно влиять на структуру знаний пользователя по изучаемому вопросу. Поэтому при построении гипертекстовых систем следует уделить внимание не только тому, как разбить исходный текст на части, но и тому, насколько пользователю будет понятно, легко и удобно работать с этими частями текста.

К основным элементам гипертекстовой технологии относятся:

• информационный фрагмент;
Рис. 6.1. Гипертекстовая форма представления информации
• тема;
• узлы;
• ссылки.

Информационный фрагмент гипертекста может представлять собой линейную последовательность строк текста, рисунок, видеофрагмент, аудиофрагмент.

Тема содержит краткое название информационного фрагмента. Информационный фрагмент может состоять целиком из множества тем либо включать в себя одну или несколько тем наряду с прочей информацией.

Узлом в гипертексте называется информационный фрагмент, из которого возможен переход к другим информационным фрагментам гипертекста.

Ссылка представляет собой слово, фразу или набор фраз, с помощью которых осуществляется переход от одного узла к другому. Ссылки могут быть референтными или организационными.

Референтные ссылки — это наиболее типичный вид ссылок в гипертекстах. Они, как правило, имеют два конца, обычно это направленные связи, хотя большинство гипертекстовых информационных систем поддерживает и обратное движение по ссылке. Исходный конец референтной ссылки называется «источник». Логически это отдельная точка или область в тексте. Другой конец называется «назначением» — это определенная точка или область в гипертексте. С источником ссылки связывается некоторая пометка, указывающая наличие ссылки, — она показывает имя ссылки, обычно изображается в виде последовательности символов и высвечивается как отдельная единица текста. Например, при щелчке по термину появится информационный фрагмент, разъясняющий значение этого термина.

Организационные ссылки устанавливают явные связи между двумя точками гипертекста и отличаются от референтных тем, что поддерживают иерархическую структуру в гипертексте. Организационные ссылки связывают узел-родитель с узлами-сыновьями и, таким образом, формируют древовидный подграф в рамках общего гипертекстового сетевого подграфа. Такие ссылки часто соответствуют отношению «быть частным случаем», и по этой причине операции над этими ссылками (при построении гипертекста) отличаются от операций над референтными ссылками [11, 18].
Кроме явных реперентных и организационных ссылок, в некоторых гипертекстовых системах имеется возможность устанавливать неявные ссылки через использование ключевых слов. Для этого гипертекстовая система должна иметь возможность сквозного поиска заданный подстроки среди всех узлов гипертекста (в некотором порядке), а в самом гипертексте должны активно использоваться ключевые слова [11]. Примером использования неявных ссылок может служить поиск в Интернет-каталогах, таких, как Yandex, Rambler, Yahoo и т. д. (рис. 6.2).

С функциональной точки зрения следование по ссылкам и поиск подстроки близки: каждая операция — это способ получить доступ к интересующему узлу. Ссылки приводят к единст-
венному узлу, а поиск по ключевому слову — к множеству узлов. В настоящее время основными группами приложений, использующих гипертекстовую технологию, являются:

1) глобальная информационная система WWW (World Wide Web — всемирная паутина);

2) справочные системы прикладных программ и операционных систем (например, в OC Windows используется два вида справочных систем: традиционная система справки, обрабатывающая файлы с расширением *.hlp и более современная справочная система HTML Help, использующая скомпилированные HTML-файлы, имеющие расширение *.chm;

3) компьютерные справочно-правовые системы (например, ГАРАНТ, КОНСУЛЬТАНТ ПЛЮС и т. д.).

6.2. Основы web-технологии


В основе создания web-документов лежит язык разметки гипертекста HTML. Аббревиатура HTML расшифровывается как HyperText Markup Language, т. е. язык разметки гипертекста. Несмотря на то, что слово «язык» присутствует в этой аббревиатуре, HTML не является языком программирования — он предназначен для разметки текстовых документов (т. е. для форматирования текста).

По формату файлы HTML представляют собой текстовые файлы, содержащие только ASCII-символы. Следовательно, для создания HTML-страниц любой сложности в принципе не требуется ничего, кроме простейшего текстового редактора (например, Блокнот). Несмотря на это, в настоящее время существует большое количество визуальных HTML-редакторов, значительн-
но облегчающих и ускоряющих процесс создания web-страниц (MS Front Page 2003, Home Site и др.).

То, как будет выглядеть текст в Web-браузере, определяют метки (или тэги, от англ. *tags*). Все, что находится между скобками *< и >* — это тэги. Текст, не находящийся между такими скобками *< >*, виден при просмотре в браузере, в то время как тэги при просмотре не отображаются.

Различают *открывающие* и *закрывающие* тэги. Признаком закрывающего тэга служит знак деления в его начале (например, *<FONT>* — открывающий тэг, а *</FONT>* — закрывающий). Текст, находящийся между открывающим и закрывающим тэгами, форматируется в соответствии с этим тэгом, например, фрагмент HTML-кода: *<B>Пример текста</B>* в браузере будет выглядеть как: Пример текста (тэг *<B>* предписывает выводить текст полужирным шрифтом).

Кроме тэгов, служащих для форматирования текста, существуют и тэги, являющиеся самостоятельными элементами страницы, например, тэг *<HR>* , выводящий на странице горизонтальную линию. Эти тэги не требуют закрывающих тэгов.

Подавляющее большинство тэгов в языке HTML позволяют задавать определенные атрибуты. Например, тэг *<FONT>* может иметь такие атрибуты, как color (определяет цвет текста), size (определяет размер шрифта) и т. д. Атрибуты прописываются внутри открывающего тэга и разделяются пробелами.

Синтаксис записи атрибутов в языке HTML выглядит следующим образом:

*<ТЭГ АТРИБУТ1*="ЗНАЧЕНИЕ" АТРИБУТ2*="ЗНАЧЕНИЕ"...>*

Например, фрагмент HTML-кода: *<FONT SIZE=*="4" FACE=*="ARIAL">Пример текста</FONT>* в браузере будет выглядеть следующим образом: Пример текста

Все документы HTML имеют одну и ту же структуру, которая выглядит следующим образом:

*<HTML>*
*<HEAD>*
*<TITLE>*Заголовок документа</TITLE>*
*</HEAD>*
*<BODY>*Тело документа</BODY>*
*</HTML>*
Тэг `<HTML>` являетсязнаком того, что данный файл содержит документ HTML. Все, что находится между тэгами `<HTML>` и `</HTML>`, браузер рассматривает как HTML-код.

Тэги `<HEAD>`-`</HEAD>` ограничивают раздел заголовка документа. Этот раздел не включает собственно содержания документа. В него входят только тэги, относящиеся к документу в целом: описание документа, ключевые слова для использования поисковыми системами и т. д. То есть информация, указанная в этом разделе, обычно предназначена не для читателей, а для роботов поисковых систем.

В этот же раздел входит заголовок документа, который задается с помощью тэга `<TITLE>`. Текст, помещенный между тэгами `<TITLE>` и `</TITLE>`, выводится в строке заголовка программы-браузера.

Тэг `<BODY>` задает основную часть документа — его «тело». Информация, размещенная между тэгами `<BODY>` и `</BODY>`, выводится в окне браузера и представляет собой содержание web-страницы.

Вышеперечисленные тэги являются обязательными и должны присутствовать в любом HTML-документе.

Файлы HTML-документов обычно имеют расширения *.html или *.htm. Встречаются также файлы с расширениями shtml, phtml, php, cgi и т. д. Файлы с такими расширениями динамически формируются web-сервером. Кириллица в именах web-документов не используется, поскольку кодировка кириллических символов различается в операционных системах Windows, где web-страницы обычно создаются и просматриваются, и в OC семейства UNIX (Free BSD, Linux), под управлением которых работает более 70 % web-серверов. Кроме того, использование кириллических символов в адресах URL может затруднить формирование HTTP-запроса к web-серверу.

Изучение языка гипертекстовой разметки HTML не входит в задачи данного курса, поскольку данный язык изучается в курсе «Программное обеспечение компьютерных сетей». Тем не менее современные редакторы позволяют создавать полноценные HTML-страницы и целые web-узлы и без знания HTML. Наиболее популярными из них являются Microsoft FrontPage, входящий в расширенные выпуски пакетов Microsoft Office, и Macromedia Dreamweaver.
Контрольные вопросы

1. Что такое гипертекстовая технология? Для чего она предназначена?
2. Назовите пять этапов, включенных в технологию построения гипертекста.
3. Перечислите основные элементы гипертекстовой технологии.
4. Что такое информационный фрагмент текста, тема, узел, ссылка?
5. В чем заключаются основные различия между референтными и организационными ссылками?
6. Назовите основные группы приложений, использующих гипертекстовую технологию.
7. Что представляет собой система World Wide Web?
8. Что такое HTML, и для чего это предназначено?
9. Что такое тэги, для чего они предназначены и какими бывают?
10. Почему кириллица не используется в именах web-документов?