



РАЗРАБОТКА И ОФОРМЛЕНИЕ ТЕКСТОВОГО СОДЕРЖАНИЯ САЙТОВ

Более 100 практических
советов и рекомендаций

Более 100 конкретных
примеров

Более 50 распространенных
ошибок разработчиков

Более 30 стандартов
и официальных документов



Ростислав Чебыкин

**РАЗРАБОТКА И ОФОРМЛЕНИЕ
ТЕКСТОВОГО
СОДЕРЖАНИЯ
САЙТОВ**

Санкт-Петербург

«БХВ-Петербург»

2004

УДК 681.3.06
ББК 32.973
Ч-34

Чебыкин Р. И.

Ч-34 Разработка и оформление текстового содержания сайтов. — СПб.: БХВ-Петербург, 2004. — 528 с.: ил.

ISBN 5-94157-396-0

Рассматриваются основные аспекты разработки текстового содержания сайтов, прежде всего русскоязычных. Излагаются правила и нормы синтаксического оформления и типографики русского текста, сведения о представлении текстовых данных в компьютере и кодировках символов. Изучается техника разметки и оформления текстового содержания веб-страниц средствами HTML и CSS. Обсуждаются принципы составления текста для сайтов, а также вопросы удобства, доступности, эффективности и привлекательности текстового содержания сайтов для пользователей.

Для веб-разработчиков

УДК 681.3.06
ББК 32.973

Группа подготовки издания:

Главный редактор	<i>Екатерина Кондукова</i>
Зам. гл. редактора	<i>Евгений Рыбаков</i>
Зав. редакцией	<i>Григорий Добин</i>
Компьютерная верстка	<i>Натальи Караваевой</i>
Корректор	<i>Виктория Пиотровская</i>
Дизайн серии	<i>Инна Тачина</i>
Дизайн обложки	<i>Игоря Цырульниковой</i>
Зав. производством	<i>Николай Тверских</i>

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 08.10.04.

Формат 70×100^{1/16}. Печать офсетная. Усл. печ. л. 42,57.

Тираж 3000 экз. Заказ №

"БХВ-Петербург", 190005, Санкт-Петербург, Измайловский пр., 29.

Гигиеническое заключение на продукцию, товар № 77.99.02.953.Д.001537.03.02 от 13.03.2002 г. выдано Департаментом ГСЭН Минздрава России.

Отпечатано с готовых диапозитивов
в ГУП "Типография "Наука"
199034, Санкт-Петербург, 9 линия, 12

ISBN 5-94157-396-0

© Чебыкин Р. И., 2004
© Оформление, издательство "БХВ-Петербург", 2004

Содержание

Добро пожаловать! Предисловие	1
Summa rei. О чем эта книга.....	1
Читатель — половина писателя. Для кого эта книга	2
Этапы большого пути. Как устроена эта книга	3
Веселые картинки. Условные обозначения.....	5
Добавить в «Избранное». Поддержка этой книги в Интернете.....	5
За все хорошее. Благодарности.....	6

ЧАСТЬ I. УРОКИ СЛОВЕСНОСТИ.

ТЕКСТ — ОСНОВНОЙ НОСИТЕЛЬ ИНФОРМАЦИИ В ИНТЕРНЕТЕ.....	7
--	----------

Глава 1. Текст и Сеть. Почему при разработке сайтов необходимо уделять особенное внимание текстовому содержанию.....	9
---	----------

1.1. Текст как он есть. Какие свойства отличают текст от других форм представления информации и делают его основным носителем информации в Интернете.....	10
1.1.1. Статичность	10
1.1.2. Дискретность	11
1.1.3. Линейность	13
1.1.4. Множество представлений.....	14
1.1.5. Удобство представления в компьютере	15
1.1.6. Текст и браузер.....	17
1.2. Параллельные миры. Чем сайты и печатные издания похожи друг на друга и чем они различаются	18
1.2.1. Материальное — информационное	19
1.2.2. Традиционное — новое	21
Гиперссылки	23
1.2.3. Линейное — иерархическое	23
1.2.4. Безответное — интерактивное.....	25
1.2.5. Управляемое — неуправляемое	26
1.2.6. Что же их объединяет?	29

Глава 2. От «А» до «Я». Важные особенности грамматики и типографики для сайтов на русском и английском языках	32
2.1. Что написано пером. Некоторые правила грамматики и типографики	33
2.1.1. Пробелы	34
2.1.2. Скобки и кавычки.....	35
2.1.3. Знаки препинания	37
Знаки препинания при скобках и кавычках.....	39
2.1.4. Дефис, тире, минус.....	40
2.1.5. Математические знаки	43
2.1.6. Числа	45
2.1.7. Специальные знаки	47
2.1.8. Сокращения	50
2.1.9. Переносы	54
2.1.10. Имена собственные, наименования	56
2.1.11. Заголовки и заголовочные надписи	62
2.1.12. Перечни (списки).....	64
2.1.13. Выделения.....	66
2.2. Новое — значит лучшее? Современные явления в сетевом русском языке	68
2.2.1. «Браузер» против «обозревателя». Откуда берутся специальные термины, связанные с Интернетом	69
Норма.....	70
Традиция	70
Русскоязычный эквивалент.....	71
Транскрипция или транслитерация	72
Ориентация на аудиторию	73
Благозвучность.....	74
2.2.2. Админ пингует хост. О специальной терминологии и сетевом жаргоне	75
2.2.3. Знаки времени. Новшества сетевой типографики	78
Новые сокращения.....	78
Типографские причуды	80
Эмотиконы.....	81
Обсценная типографика	83
2.2.4. Смесь французского с нижегородским. О включении в текст на русском языке слов, записанных на других языках.....	84
 ЧАСТЬ II. ДЕЛО МАСТЕРА БОИТСЯ. ТЕХНИКА РАЗРАБОТКИ И КОДИРОВАНИЯ ТЕКСТОВОГО СОДЕРЖАНИЯ САЙТОВ	89
Глава 3. Двоичная сюита. Представление текста в компьютере: кодировки символов, форматы текстовых файлов	91
3.1. Азбука для компьютера. Кодировки символов	91
3.1.1. Кодировки и кодовые таблицы	92

3.1.2. Двоичная и шестнадцатеричная системы счисления	95
3.1.3. ASCII	99
Перевод строки	105
Табуляция	107
3.1.4. Расширенные кодировки ASCII	110
Неразрывный пробел и мягкий перенос	114
3.1.5. Кириллица в расширенных кодировках ASCII	115
Проблемы перекодировки	122
3.1.6. Unicode — решение многих проблем	125
ISO 10646 и Unicode	126
UTF-8 и другие UTF	132
Трудности с использованием Unicode на практике	135
3.2. Лики текста. Форматы текстовых файлов	138
3.2.1. Двоичный и текстовый режимы	138
3.2.2. Обычный текст	143
3.2.3. Форматированный текст	150
RTF (Rich Text Format)	150
DOC (Документ Microsoft Word)	152
TEX	154
HTML (HyperText Markup Language)	156
PostScript	158
PDF (Portable Document Format)	159
3.3. При всем богатстве выбора. Программы для работы с текстовыми файлами	161
3.3.1. Текстовые редакторы	163
3.3.2. Текстовые процессоры	165
WYSIWYG	167
Среды разработки сайтов	169
3.3.3. Издательские системы	172

Глава 4. «Блокнот» мой — друг мой. Эффективная работа с текстовым редактором

175

4.1. Кто на свете всех милее? Как выбрать текстовый редактор	176
4.2. От частного к общему. Свойства и возможности текстовых редакторов	182
4.2.1. Текстовый и графический режимы	182
4.2.2. Курсор	186
4.2.3. Набор и редактирование текста	190
4.2.4. Выделение текста	195
4.2.5. Буфер обмена	198
4.2.6. Символы и шрифты	200
4.2.7. Прокрутка и переводы строк	205
4.2.8. Операции с файлами	207
4.2.9. Обработка текста	211

Глава 5. Наука побеждать. Техника разметки и оформления текста с помощью HTML и CSS	214
5.1. С чего начинается Родина. HTML — язык структурной разметки гипертекста	215
5.1.1. Как это понимать? Что означают слова «структурный», «разметка» и «гипертекст»	216
5.1.2. Во первых строках письма... Шаблон документа HTML	220
Пролог документа и версии HTML	221
Валидатор и стандарты W3C	227
Тэги, элементы, атрибуты	231
5.1.3. Уроки хороших манер. Стилль кодирования	234
Основные элементы	237
Вложенные элементы	240
Комментарии	243
Пробельные символы	245
Имена файлов HTML	246
5.1.4. Строительный материал. Структурные элементы HTML	248
Строчные и блочные элементы	248
<i>title</i> — самый главный элемент	250
Заголовки: <i>h1</i> — <i>h6</i>	256
Текст: <i>p</i> и <i>pre</i>	260
Внутритекстовые выделения: <i>em</i> и <i>strong</i>	263
Другие внутритекстовые выделения	265
Блочные цитаты: <i>blockquote</i>	268
Перечни: <i>ul</i> , <i>ol</i> , <i>dl</i>	270
Таблицы	278
Гиперссылки: элемент <i>a</i>	284
Изображения и их текстовые эквиваленты	296
5.1.5. Полезные мелочи. Подстановки символов	301
Обязательные подстановки	306
Неразрывный пробел	307
Специальное использование подстановок	309
5.1.6. Что дальше? Перспективы развития HTML и XHTML	312
5.2. Красота требует жертв. Оформление текстового содержания веб-страниц с помощью CSS	314
5.2.1. Три кита. Содержание, структура, представление	315
5.2.2. Курс молодого бойца. Основы кодирования таблиц стилей	317
Начальные сведения	317
Замечания о кодировании CSS	320
Виды селекторов CSS	324
Преимущества и недостатки CSS	327
5.2.3. Чтоб картинка ожила. Свойства CSS для оформления текстового содержания веб-страницы	329
Свойства шрифта	330
Свойства блоков текста	344

ЧАСТЬ III. ВСЕ ДЛЯ ЧЕЛОВЕКА. КАК ДОБИТЬСЯ, ЧТОБЫ ТЕКСТОВОЕ СОДЕРЖАНИЕ САЙТА СООТВЕТСТВОВАЛО ИНТЕРЕСАМ И ЗАДАЧАМ ПОЛЬЗОВАТЕЛЕЙ.....363

Глава 6. Ради красного словца. Как составлять текст для размещения на сайте.....365

- 6.1. Доброе слово и кошке приятно. Как составить текст,
производящий на пользователя нужное впечатление..... 366
 - 6.1.1. Совершенство малых форм. Самые короткие надписи..... 366
 - 6.1.2. Туда-сюда. Особенности составления текста для навигации 371
 - 6.1.3. Короче! Соблюдайте краткость в тексте на сайтах..... 377
 - 6.1.4. Проще простого. Пишите простой текст..... 382
 - 6.1.5. Сверим наши стили. Придерживайтесь единого стиля..... 386
 - 6.1.6. И смех и грех. Не злоупотребляйте метафорами и юмором..... 388
 - 6.1.7. Честно говоря. Размещайте на сайте только достоверную
информацию 391
 - 6.1.8. Внесите разнообразие! Не пренебрегайте средствами
оформления..... 394
 - 6.1.9. Заставьте текст работать! Применяйте элементы,
с которыми пользователь может взаимодействовать..... 396
 - 6.1.10. В особо крупных размерах. Что делать
с крупными текстами..... 399
 - 6.1.11. Вы ничего не забыли? Еще несколько
распространенных ошибок..... 403
 - Время и дата..... 403
 - Контактная информация..... 404
 - Under construction..... 405
- 6.2. Не забывайте про роботов. Как оптимизировать
текстовое содержание сайта для поисковых систем 406
 - 6.2.1. Облик Искателей. Общая информация
о поисковых системах..... 406
 - 6.2.2. Разметка превыше всего. Грамотно употребляйте
структурную разметку 409
 - 6.2.3. Чтоб язык не заплетался. Учитывайте особенности
русского языка..... 412
 - 6.2.4. Хитрости полезные и вредные. Как еще повисить
«находимость» вашего сайта..... 415

Глава 7. Пользователь всегда прав. Удобство, доступность, эффективность и привлекательность текстового содержания сайта.....419

- 7.1. Чтоб никто не ушел обиженным. Основные принципы удобства
текстового содержания 422

7.1.1. Привычка — вторая натура. Учитывайте естественное поведение и привычки пользователей	422
7.1.2. Лень — двигатель прогресса. Пользователь не должен прилагать лишних усилий	428
Копирование, сохранение и печать текста	437
7.1.3. Доверяй, но проверяй. О тестировании удобства сайта	442
Редактор моей мечты. Поиск подходящего текстового редактора на сайте <i>www.listsoft.com</i>	445
Belle! Поиск нот популярного трио из «Notre-Dame de Paris»	452
7.1.4. Как вам будет угодно. Персонализация сайтов	463
7.2. Это доступно? Критерии доступности текстового содержания сайтов	466
7.2.1. Найдите десять отличий. «Машинный фактор» доступности	467
Устройства просмотра	467
Операционная система	469
Видеорежим	474
Браузер	476
7.2.2. Esse homo. «Человеческий фактор» доступности	485
7.2.3. Скрижали Закона. Принципы доступности сайтов от Консорциума W3C	489
Стандарты	489
Структура информации	491
Разделение структуры и представления	493
Оформление	496
Персонализация	497
Степени соответствия WCAG 1.0	498
7.3. ...И люди к вам потянутся. Замечания о привлекательности сайта	499
7.3.1. Содержание	499
7.3.2. Структура и навигация	501
7.3.3. Оформление	503
Пора и честь знать. Заключение	507
Смотрит в книгу — видит... Рекомендуемая литература	513
Орфография, пунктуация и стилистика русского языка	513
Издательское дело	514
Разработка сайтов	517
Ресурсы Всемирной Паутины	518
Стандарты и документация	518
Другие ресурсы	519
Браузеры	519

Добро пожаловать!

Предисловие

Когда выходишь на эстраду,
Стремиться нужно к одному:
Всем рассказать немедля надо,
Кто ты, зачем и почему.

Л. Зорин. Покровские ворота

1.7.14. An abstract **should** briefly describe the document. The word “Abstract” **must** be marked up with an h2 element.

World Wide Web Consortium Publication Rules¹



Внимание!

Эта книга посвящена всем аспектам разработки текстового содержания сайтов в Интернете, предназначена для составителей сетевого текста, редакторов, корректоров и кодировщиков, а также может быть интересна всем, чья деятельность связана с созданием сайтов.

Страница поддержки книги в Интернете — www.philigon.ru/books/webtext.

Summa rei²

О чем эта книга

— Расскажите кратенько, в двух словах, о чем ваша книга.

— Эх, если бы я мог рассказать об этом в двух словах, я бы не стал писать целую книгу.

Литературный анекдот

Эта книга — обо всем, что связано с разработкой текстового содержания сайтов в Интернете.

¹ 1.7.14. В обзоре следует кратко описывать документ. Слово «Обзор» должно быть размечено элементом h2. (*Правила публикаций Консорциума W3C*)

² Суть дела (*лат.*).

Вот наиболее важные темы книги:

- ❑ чем сайты и печатные издания похожи и чем различаются; какие практические выводы следуют из этих сходств и различий;
- ❑ какие правила русской и английской грамматики и типографики особенно важно учитывать при разработке сайтов;
- ❑ что нужно знать разработчику сайтов о технической стороне представления текста в компьютере, о кодировках и форматах текстовых файлов;
- ❑ как экономить время и повысить эффективность своей работы с текстовым редактором и другими программами, используемыми при разработке сайтов;
- ❑ как правильно и грамотно применять языки HTML и CSS для структурной разметки и оформления текстового содержания сайтов;
- ❑ как составлять текст, который сделает сайт эффективнее и поможет привлечь и удержать пользователей;
- ❑ как улучшить удобство и доступность текстового содержания сайта.

Читатель — половина писателя

Для кого эта книга

Не давайте святыни псам и не бросайте жемчуга вашего перед свиньями, чтоб они не попрали его ногами своими и, обратившись, не растерзали вас.

Евангелие от Матфея, гл. 7, ст. 6

Книга, в первую очередь, предназначена для двух категорий читателей.

Во-первых, для всех, кто разрабатывает или собирается разрабатывать текстовое содержание сайтов:

- ❑ авторам сетевого текста, редакторам, корректорам и верстальщикам сайтов она поможет работать более продуктивно, избегать досадных ошибок и делать сайты привлекательнее для пользователей;
- ❑ всем, кто учится созданию сайтов, — даст представление о разработке текстового содержания сайтов и позволит принять осознанное решение, стоит ли делать это занятие областью своей специализации.

Во-вторых, эта книга для всех, кому приходится оценивать качество сайтов:

- ❑ руководителям сетевых проектов она поможет организовывать деятельность, связанную с разработкой текстов, а также контролировать работу авторов, редакторов и верстальщиков;
- ❑ экспертам в области удобства и доступности сайтов — оценивать удобство и доступность текстового содержания.

Автор надеется, что читатели знакомы с основами компьютерной грамотности, что они хотя бы в общих чертах представляют себе, что такое Интернет, и имеют опыт посещения сайтов. Для чтения этой книги больше ничего не потребуется. Хотя более понятна она будет для тех, кто уже занимался разработкой сайтов, — хотя бы на любительском уровне — и имеет представление о HTML и CSS.



Примечание

Ростислав Чебыкин (р. 1977) — преподаватель информационных технологий в Экономическом бизнес-колледже (Москва) и международной школе бизнеса «Орион-Интер». Эта книга основана на одном из разделов курса «Веб-мастер», который автор ведет в течение нескольких лет. Персональный сайт автора — www.philigon.ru. Адрес электронной почты: philigon@master.elserv.msk.su.

Этапы большого пути

Как устроена эта книга

Рубрикация — это система заголовков издания и возглавляемых ими подразделов (рубрик), выражающая логическую или иную связь и соподчиненность обозначаемых ими частей текста (произведений, разделов, глав, параграфов).

*А. Э. Мильчин, Л. К. Чельцова.
Справочник издателя и автора*

Книга состоит из трех частей.

Часть 1 — «Уроки словесности» — это теоретический базис последующего изложения. Здесь читатель узнает:

- что такое текст, каковы его характерные особенности в сравнении с другими формами представления информации;
- почему текст является основным носителем информации в Интернете;
- какие сходства и различия между сайтами и печатными изданиями должен учитывать разработчик;
- как избегать наиболее распространенных грамматических ошибок в сетевом тексте (прописные и строчные буквы, сокращения, заголовки и т. д.);
- как грамотно использовать в тексте специальные типографские символы (тире, кавычки, пробелы, многоточие и т. д.);

- ❑ стоит ли при составлении текста для сайтов употреблять текстовые новшества (смайлики, жаргон, иностранные слова и т. д.).

Часть 2 — «**Дело мастера боится**» — посвящена практической стороне разработки текстового содержания сайтов. Здесь рассказывается:

- ❑ как текст хранится и преобразуется в компьютере, какие программные средства применяются для работы с ним;
- ❑ что такое кодировки символов и зачем разработчику сайтов нужно в них разбираться;
- ❑ как повысить эффективность своей работы с текстовым редактором и другими программами разработки;
- ❑ как правильно использовать язык HTML для структурной разметки текстового содержания;
- ❑ как оформлять текстовые элементы сайта с помощью языка CSS.

Часть 3 — «**Все для человека**» — касается вопросов удобства и доступности текстового содержания сайта для пользователей. Здесь освещены следующие темы:

- ❑ как составлять эффективный текст, предназначенный для размещения на сайте;
- ❑ как сделать так, чтобы пользователям было удобно работать с текстом на сайте;
- ❑ как повысить доступность текстового содержания для различных категорий посетителей, включая пользователей с ограниченными возможностями;
- ❑ как текстовое содержание поможет привлечь пользователей и удержать их на сайте;
- ❑ как обеспечить единство текста с другими компонентами сайта.

Книга построена как учебник, но ее можно использовать и как справочник. Чтобы читатель мог легко найти нужную информацию, в конце книги есть предметный указатель.



Примечание

Основная часть книги написана в середине 2004 года, и в ней отражена практика, сложившаяся к этому моменту. Читая книгу, не забывайте, что жизнь не стоит на месте.

Веселые картинки

Условные обозначения

Иероглифы представляют собой идеограммы, то есть обозначают целые понятия. Такая форма фиксации информации оказала значительное влияние на характер мышления людей, способствовала формированию структурно-образного восприятия.

Япония от «А» до «Я»



Такой символ (U+26A0) обозначает краткий конспект последующего текста или наиболее важные выводы, на которые рекомендуется обратить особое внимание. Если вам не хочется или некогда читать книгу целиком — прочтите хотя бы то, что помечено этим символом.



Таким символом (U+2611) обозначаются полезные замечания и отступления. Иногда считают, будто подобные замечания появляются потому, что автор очень хотел что-то сказать, но никак не смог уместить это в основном тексте. А иногда оказывается, что без замечаний основной текст не имеет никакого смысла.



Такой символ (U+7EB8, «китайская грамота») обозначает необязательный материал, который можно пропустить при первом чтении книги. Возможно, его даже нужно пропускать: ведь тогда появится стимул перечитать книгу снова!

Добавить в «Избранное»

Поддержка этой книги в Интернете

Один из недостатков обычного издания книги в том, что невозможно мгновенно перепечатать и распространить книгу заново, если появляется необходимость исправить ошибки или внести какие-либо технологические изменения. В Сети же, наоборот, можно мгновенно изменить любую информацию. Для этого я создал сайт, на котором публикую все исправления.

Я. Нильсен. Веб-дизайн

Этой книге посвящена страница www.philigon.ru/books/webtext. Там находится список обнаруженных ошибок и опечаток, избранные фрагменты книги, рекомендации для преподавателей и некоторая другая информация.



Примечание

Один из недостатков размещения информации в Интернете — нестабильность и недолговечность. Может быть, сейчас, когда вы держите в руках эту книгу, давно уже нет ни автора, ни его сайта, ни самого Интернета — так что ссылка на страницу поддержки неактуальна. Поэтому автору остается только обещать позаботиться о том, чтобы эта ссылка проработала как можно дольше и чтобы ее исчезновение (которое, увы, рано или поздно наступит) не принесло никому неудобства.

За все хорошее Благодарности

Ложки у меня пациенты много раз глотали, не скрою, но вот чтобы так — за обедом на десерт и острый предмет — замечательно. За это вам наша искренняя сердечная благодарность.

Г. Горин. Формула любви

Автор благодарен многим людям, которые, сами того не подозревая, оказали неоценимую помощь в работе над этой книгой. Вот лишь некоторые из них: Тим Бернерс-Ли, Людвиг ван Бетховен, Дмитрий Гайдук, Дональд Кнут, Льюис Кэрролл, Леонид Лиходеев, Якоб Нильсен, Чарльз Петцольд, Александр Привалов, Сергей Рахманинов, Владимир Успенский, Джеймс Феличи, Франциск Ассизский, Даниил Хармс, Даглас Хофштадтер, Ян Чихольд, Морис Эшер.

```
<?xml version= 1.0
<!DOCTYPE html PUBLIC
"-//W3C//DTD XHTML
"http://www.w3.org
<html xmlns="http://
lang="ru" xml:lang
<!-- Редактировал:
<head>
<title>Евгений Он
<link rel="styles
href="index.css
</head>
```

Часть I

Уроки словесности

Текст — основной носитель информации в Интернете

— Что ж, начинать? — Начнем, пожалуй.

А. С. Пушкин. Евгений Онегин

ТЕКСТ (от лат. *textus* — ткань, соедине-
ние),

1) последовательность предложений, слов
(в семиотике — знаков), построенная со-
гласно правилам данного языка, данной зна-
ковой системы и образующая сообщение.

2) Словесное произведение; в художест-
венной литературе — законченное произ-
ведение либо его фрагмент, составленный
из знаков естественного языка (слов)
и сложных эстетических знаков (слагае-
мых поэтического языка, сюжета, компо-
зиции и т. д.).

3) Авторское сочинение без комментариев
и приложений к нему.

4) В издании — часть произведения, вы-
раженная в словесной форме, в отличие
от рисунков, чертежей, формул и т. п.

5) Типографский шрифт, кегль (размер)
которого равен 20 пунктам (около 7,5 мм).

Российский энциклопедический словарь

Глава 1

```
<?xml version= 1.0
<!DOCTYPE html PUBL
"-//w3c//DTD XHTML
"http://www.w3.org
<html xmlns="http://
lang="ru" xml:lang
<!-- Редактировал:
<head>
<title>Евгений Он
<link rel="styles
href="index.css
</head>
```

Текст и Сеть

Почему при разработке сайтов необходимо уделять особенное внимание текстовому содержанию

Если рассматривать «текст» в самом широком смысле этого слова, то, вероятно, 90 % всего, что мы делаем, на 90 % состоит из обработки текста.

*Л. Уолл, Т. Кристиансен, Дж. Орвант.
Программирование на Perl*

Сейчас мы знаем, что текст состоит не из линии слов, воплощающих единственное «теологическое» значение («сообщение» Автора-Бога), но из многомерного пространства, где обручаются и вступают в распрю всевозможные виды письма, из которых ни один не является изначальным: текст есть производство цитаций, исходящих из тысяч источников культуры.

Р. Барт. Смерть автора

Мы не можем представить себе сайт без текста. Бывают сайты без иллюстраций, без художественного оформления, без специальных объектов вроде фоновой музыки или анимационных роликов, но сайт без текста — это нон-сенс. Текст — основная форма представления информации в Интернете, поэтому его разработка заслуживает самого пристального внимания.

1.1. Текст как он есть

Какие свойства отличают текст от других форм представления информации и делают его основным носителем информации в Интернете

Иными словами, откажемся пока от представления о тексте как об отформатированных абзацах на двумерной печатной странице. Будем считать его одномерным потоком букв, цифр, знаков препинания и, возможно, дополнительных символов, указывающих на окончание абзаца.

Ч. Петцольд. Код



Внимание!

Основные свойства текста:

- статичность;
- дискретность;
- линейность;
- множество представлений;
- удобство представления в компьютере.

1.1.1. Статичность

Статичность — это так прекрасно! Так непостижимо величественно!

А. Поникаровская. Промежуточный этаж

Статичность текста означает, что он *не развивается во времени*. Это просто понять, если сравнить текст с музыкальной композицией или видеозаписью. Такие формы представления информации имеют определенную временную продолжительность. А читатель текста сам определяет, с какой скоростью он будет его читать и сколько времени потратит на чтение.

Статичность — одна из причин того, почему многим людям легче читать текст на иностранном языке, чем слушать этот же текст в устном изложении.

Статичность текста нарушается, если он включается в титры, бегущую строку или другие динамические элементы. Это является вынужденной мерой в кино или на рекламных конструкциях, но на сайтах нарушение статичности текста чаще всего ничем не оправдано.

**Внимание!**

Разрабатывая сайт, не нарушайте статичности текста. Пусть текст на веб-странице будет неподвижным. Не надо бегущих строк, мигающих букв и прочих динамических эффектов. Не мешайте пользователю читать текст с такой скоростью, с какой ему удобнее.

1.1.2. Дискретность

Атомы Демокрита различаются по форме и по величине; двигаясь в пустоте, они соединяются («сцепляются») между собой в силу различия по форме: у Демокрита есть атомы круглые, пирамидальные, кривые, заостренные, даже «с крючками». Так из них образуются тела, доступные нашему восприятию.

И. Фролов. Введение в философию

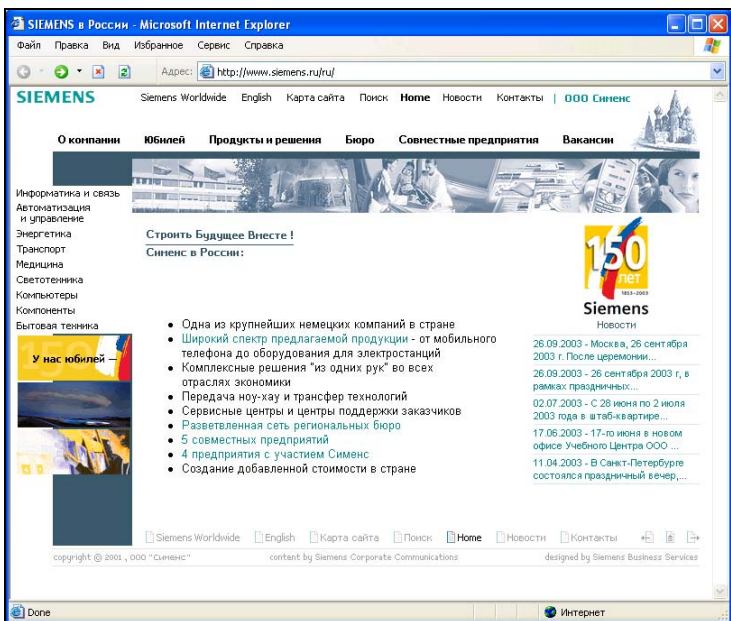
Дискретность текста означает, что он состоит из отдельных элементов — *символов*, каждый из которых рассматривается как единое целое. Этим свойством не обладают, например, изображение или звуковая запись — они являются непрерывными.

Дискретность текста на веб-странице нарушается, если он представлен не в виде последовательности символов, а в виде изображения — например, графического заголовка или панели навигации (рис. 1.1, *а*). Это вызывает сразу несколько неприятных явлений (рис. 1.1, *б*):

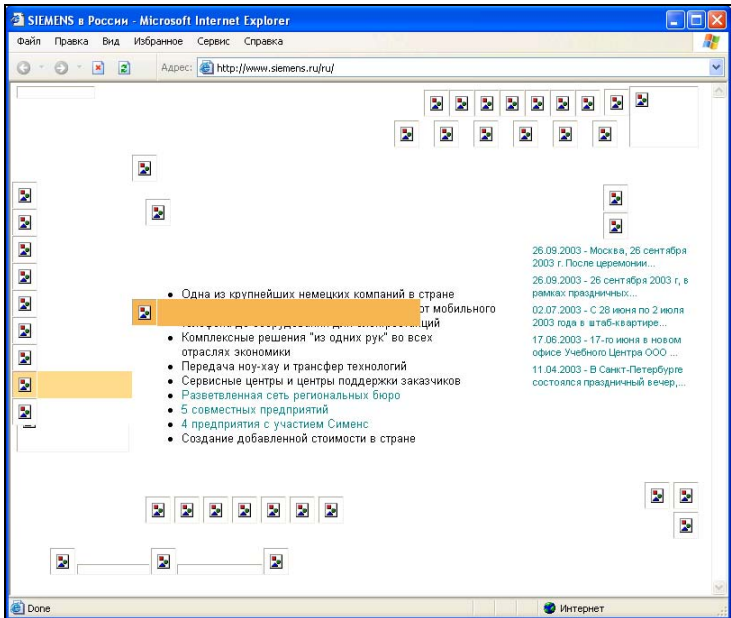
- если в браузере пользователя отключены или не поддерживаются изображения, он не сможет прочесть такой текст;
- если этот текст слишком мелкий или слишком крупный для пользователя, он не сможет увеличить или уменьшить его стандартными средствами браузера;
- если пользователю потребуется сохранить у себя такой текст, он не сможет просто скопировать его в буфер обмена и вставить в свой документ; текст придется перепечатывать вручную;
- поисковые машины не распознают текст, заключенный в изображение, и не индексируют его.

**Внимание!**

Не нарушайте дискретности текста. Убедитесь, что вся текстовая информация, которая есть на вашем сайте (включая заголовки и средства навигации), представлена именно в форме текста, а не графики.



а



б

Рис. 1.1. www.siemens.ru: если в браузере отключена загрузка изображений, панель навигации на нем становится недоступной. Надо сказать, что таких сайтов с графической навигацией со временем остается все меньше. Еще два года назад этим недостатком страдал почти каждый второй сайт, а теперь автору пришлось некоторое время потрудиться в поисках



Примечание

Все правила и рекомендации, которые приводятся в этой книге, носят самый общий характер. На практике иногда встречаются ситуации, когда правила можно и даже нужно нарушать. Но делать это следует, только если вы хорошо представляете себе, к каким неприятностям приведет нарушение правил, и твердо уверены, что в данном случае эти неприятности допустимы.



Дополнительно

Дискретность текста проявляется на нескольких уровнях. Достаточно опытный читатель воспринимает текст не по отдельным буквам, а целыми блоками. Обычно такими блоками являются слова, знакомые словосочетания и даже целые фразы. Читатель «схватывает» слово или словосочетание целиком, мысленно не разделяя его на более мелкие составляющие. Если в тексте много незнакомых сложных слов, используются громоздкие грамматические и стилистические конструкции, — это затрудняет чтение, нарушая дискретность на данном уровне. Не допускайте таких ошибок, составляя текст для своего сайта.

1.1.3. Линейность

Наиболее простой способ объединить или связать некоторое множество элементов — это «вытянуть их в линию», организовать список или очередь.

Н. Вирт. Алгоритмы и структуры данных

Линейность текста означает, что он представляет собой *одномерную* структуру, последовательность символов. Символы идут строго один за другим, что отличает текст от изображения (двумерная структура) или скульптуры (трехмерная структура).

Именно линейностью объясняется тот факт, что люди научились передавать текстовые сообщения по линиям связи гораздо раньше, чем информацию любого другого характера. Флажковая азбука, азбука Морзе, язык тамтамов — все это средства передачи линейной, последовательной информации.

Если текст заключается в изображение, нарушается не только его дискретность, но и линейность. Это еще одна причина не делать «графического» текста.

1.1.4. Множество представлений

Когда женщина стареет, с ней могут произойти многие неприятности: могут выпасть зубы, поседеть и поредеть волосы, развиться одышка, может нагрязнеть тучность, может одолеть крайняя худоба, но голос у нее не изменится. Он останется таким же, каким был у нее гимназисткой, невестой или любовницей молодого повесы.

И. Ильф, Е. Петров. Двенадцать стульев

Один и тот же текст может быть представлен множеством разных способов. Например:

Мой дядя самых честных правил.

Мой дядя самых честных правил.

Moy djadja samykh chestnykh pravil.

Несмотря на то, что внешне эти строчки заметно отличаются друг от друга (особенно третья), мы видим, что в них записан один и тот же текст. Это позволяет рассматривать сам текст отдельно от его представления, внешнего вида.

纸 **Дополнительно**

Например, вы переписываетесь с кем-то по электронной почте. В вашей почтовой программе может быть установлен шрифт Andale Mono, потому что он кажется вам наиболее приятным для глаз и удобным при наборе. А почтовая программа вашего корреспондента настроена так, чтобы отображать сообщения шрифтом Helvetica — потому что так, в свою очередь, ему удобнее читать. Это не вызывает никаких проблем — вы набираете сообщение тем шрифтом, каким удобнее вам, ваш корреспондент читает его тем шрифтом, каким удобнее ему, — а текст при этом остается тем же самым, и оба участника переписки довольны.

Пользователи Интернета, в отличие от читателей печатных изданий, имеют счастливую возможность настраивать внешний вид текста на веб-страницах таким образом, чтобы им было удобнее читать. Настройки браузеров позволяют установить для текста определенный шрифт, цвет, размер и другие параметры.



Внимание!

Не мешайте пользователям настраивать «под себя» внешний вид текста на сайтах. Убедитесь, что на вашем сайте срабатывают все браузерные средства настройки внешнего вида.

Это, кстати, еще один аргумент против «графического» текста. Если текст заключен в изображение, то пользователь не сможет изменить его параметры.

纸 *Дополнительно*

У текста немало и других свойств, отличающих его от прочих форм представления информации. Например, текст обладает *избыточностью*. Это легко продемонстрировать на следующем примере:

Слв «ткст» знчт ткнь, сцплн, сднн... В лбм слч ткст прдствлт сбй бьдннн п смсл пслдвтльнст знквх днц, снвнм свйствем ктрй влтс свзннсть, цльнсть.¹

Из этого текста удалены все гласные буквы — но его все равно можно понять и восстановить. Это и означает, что полная версия текста — избыточна.

Избыточность текста приводит к двум важным качествам:

- Текст обладает хорошей устойчивостью по отношению к потерям, помехам и искажениям, поэтому его проще передавать по линиям связи. Как видно из приведенного примера, даже если из текста при передаче пропала половина букв, он, тем не менее, поддается восстановлению. А представьте, что из музыкальной мелодии пропала половина нот — как вы будете ее восстанавливать?
- Текст хорошо сжимается без потерь. Программы компрессии иногда позволяют сжать текст в 10 раз, что для других форм представления информации невозможно. Сжатый текст занимает меньше места в устройствах хранения и, что еще более важно, быстрее передается по линиям связи.

1.1.5. Удобство представления в компьютере

Другой аспект, который является общим почти для всех приложений, заключается в том, что при их использовании требуется вводить какой-то текст. (Даже в играх вам иногда приходится вводить собственное имя в случае выигрыша.)

Дж. Раскин. Интерфейс



Внимание!

Удобство представления текста в компьютере:

- простота кодирования;
- точность представления;
- удобство хранения на носителях и передачи по каналам связи;
- удобство операций с текстом.

¹ Н. С. Валгина. Теория текста. М., 2003. С. 12. — *Авт.*

Обладая перечисленными свойствами, текст является самой удобной формой информации для представления в компьютере. Это связано со следующими факторами:

простотой кодирования — из-за дискретности текста его очень просто преобразовывать в двоичный код и обратно: достаточно поставить каждому символу в соответствие код из определенной кодовой таблицы и пользоваться этой таблицей при преобразованиях.

Это даже проще, чем представление числовой информации. Чтобы сохранить в компьютере число, над ним в большинстве случаев приходится производить вспомогательные арифметические операции (например, преобразование в дополнительный код). Кодирование изображений и звуковой информации происходит еще сложнее;

- **точностью представления.** Текст — единственная форма информации, которая может быть сохранена в компьютере с абсолютной точностью. Если снова вспомнить про числа, то даже такое незамысловатое число, как $\frac{1}{3}$, при сохранении в компьютере придется округлить до некоторого конечного количества знаков после запятой, а при сохранении изображений и звуковой информации потери еще значительнее. При сохранении текста не теряется *ничего*;
- **удобством хранения на носителях и передачи по каналам связи** — чтобы сохранить информацию в оперативной памяти компьютера, на жестком диске или на любом другом носителе или чтобы передать ее (как между устройствами внутри одного компьютера, так и по линиям связи между разными компьютерами), она должна быть преобразована в линейную последовательность байтов. А текст не надо никуда преобразовывать, он и так линейный;
- **удобством операций с текстом** — обрабатывать текст на компьютере гораздо проще, чем любую другую информацию. Например, из текста легко можно удалить отдельный символ, слово или целое предложение, и текст после этого будет выглядеть так, как будто этого удаленного фрагмента вообще не было. А, скажем, удалить фигуру человека из групповой фотографии так, чтобы фотография после этого казалась нетронутой, — непростая задача, которая иногда не под силу даже профессионалам компьютерной графики;
- **другими факторами** — кроме перечисленных, можно назвать еще множество факторов удобства представления текста в компьютере. Например, работа с текстом требует гораздо меньше вычислительных ресурсов, чем работа с другими формами представления информации — скажем, с изображениями или со звуком, не говоря уже о видео. Текст предъявляет меньше требований к квалификации разработчика. В самом деле, вы легко можете в точности перепечатать на компьютере тот самый текст, который вы сейчас читаете. А можете ли вы так же легко в точности срисовать рис. 1.1?

纸 *Дополнительно*

У текста есть и недостатки по сравнению с другими формами представления информации. Например, чтобы прочесть текст, нужно знать язык, на котором он написан. Изображения или музыкальные записи не имеют этого недостатка.

С этим связаны споры о том, целесообразно ли использовать на сайтах пиктограммы: например, символическое изображение домика в качестве ссылки на главную страницу или конверта в качестве ссылки на адрес электронной почты. Сторонники пиктограмм утверждают, что такие картинки будут понятны даже тем пользователям, которые не могут прочесть текст на сайте. Противники пиктограмм считают, что они сбивают с толку: например, пользователь может подумать, что изображение домика ведет на страницу, посвященную строительству или домашнему хозяйству.

Впрочем, здесь нетрудно достичь компромисса: можно помещать на страницах пиктограммы и сопровождать их поясняющим текстом.



Примечание

Человеку свойственно ошибаться, и авторы книг — не исключение из этого правила. Какой бы хорошей ни была книга, в ней всегда найдутся какие-то недостатки, всегда обнаружится, что можно исправить и улучшить. Если вы считаете, что автор этой книги в чем-то неправ, — что же, это вполне возможно. Пожалуйста, в этом случае напишите автору по адресу: philigon@master.elserv.msk.su. Ведь автор сможет исправить свои ошибки, только если узнает о них!

1.1.6. Текст и браузер

This guideline emphasizes the importance of providing text equivalents of non-text content (images, pre-recorded audio, video). The power of text equivalents lies in their capacity to be rendered in ways that are accessible to people from various disability groups using a variety of technologies.

Web Content Accessibility Guidelines 1.0²

Существует огромное количество разнообразных браузеров, с помощью которых пользователи просматривают сайты. В этих браузерах могут не под-

² Этот принцип говорит о том, как важно предоставлять текстовые эквиваленты для нетекстового содержания (изображений, аудио- и видеозаписей). Преимущество текстовых эквивалентов в том, что они могут отображаться теми способами, которые доступны для пользователей с ограниченными возможностями, применяющих самые разные технологии доступа к сайтам. (*Принципы доступности содержания сайтов, редакция 1.0*)

держиваться или быть отключены изображения, анимационные ролики, звук, JavaScript, апплеты, таблицы стилей и все остальное. Но текст — это единственное, что поддерживается всегда и в любом браузере.

Таким образом, вы как разработчик можете быть уверены, что ваш текст в точности дойдет до пользователя, независимо от того, чем он смотрит ваш сайт. А для других составляющих сайта этого гарантировать нельзя.

Поэтому все нетекстовые компоненты информационного содержания сайта (например, изображения, анимационные ролики или апплеты) рекомендует-ся сопровождать альтернативным текстом, который должен увидеть пользо-ватель, если сам нетекстовый объект будет для него недоступен.

1.2. Параллельные миры

Чем сайты и печатные издания похожи друг на друга и чем они различаются

...И плотная сеть паутины
Упала, окутала ложе...

А. Ахматова

...Можно прийти к выводу, что техноло-гии производства аппаратного обеспече-ния — это сдерживающий фактор, и что только к 2007 году информация, доступ-ная через Сеть, целиком заменит обычные книги. Издатели, будьте бдительны! Это точно произойдет.

Я. Нильсен. Веб-дизайн

Есть мнение, что между размещением информации на сайте и публикаци-ей в печатных изданиях нет никакой принципиальной разницы: просто в одном случае пользователь просматривает документ на экране компью-тера, а в другом — на листе бумаги. Поэтому все принципы и правила, относящиеся к традиционной печати, справедливы и в Сети. Но есть и про-тивоположное мнение: Сеть — это принципиально новый носитель ин-формации, поэтому к ней неприменимы принципы бумажных публикаций.

Как это часто бывает, истина находится где-то посередине. Чтобы добиться успеха в создании сайтов, нужно хорошо представлять себе, чем они похожи на печатные издания и чем отличаются.

**Внимание!****Печатное издание**

Материальный объект

Первые книги — VIII век до н. э.

Линейная структура

Статичное восприятие читателем

Полностью контролируется издателем

Сайт

Информационный объект

Первые сайты — конец 1980-х

Иерархическая структура

Интерактивное взаимодействие с пользователем

В существенной степени контролируется пользователем

1.2.1. Материальное — информационное

Я люблю чувствовать, осязать, а в Интернете я не чувствую... Вот вас я могу потрогать, а Интернет — не могу.

Б. Гребенщиков (из интервью)

Печатные издания. Каждый экземпляр печатного издания представляет собой материальный объект: его можно потрогать, взять в руки, перенести с места на место, — наконец, уничтожить.

Сайты. Сайт — информационный объект. Это значит, что он не связан ни с каким определенным материальным носителем так прочно, как печатный текст. Работоспособность одного сайта может обеспечиваться десятками и даже сотнями различных серверов, разбросанных по всему миру. Одновременно с этим сайт существует в виде электрических сигналов, передаваемых по многочисленным линиям связи, а также в виде страниц, которые отображаются на компьютерах конечных пользователей. Сайт — это скорее не физический предмет, а служба, услуга. В этом смысле сайт похож не на печатные издания, а на такие способы распространения информации, как радио или телевидение.

Сайты обладают всеми свойствами информационных объектов:

- легко распространяются** — как только веб-страница появляется на общедоступном сервере, ее в ту же минуту можно загрузить на миллионах компьютеров по всему земному шару. Количество просмотров страницы практически не ограничено;
- легко копируются** — вы всегда можете сохранить веб-страницу на дискете и передать ее своему приятелю, у которого нет доступа в Интернет. Можно гарантировать, что ваш приятель получит в точности ту же страницу, которая до этого была у вас;

纸 *Дополнительно*

В некоторых источниках можно встретить такое определение: «Если в процессе взаимодействия двух объектов один объект передает другому некую субстанцию, а сам при этом этой субстанции не лишается, — то такое взаимодействие называется информационным, а эта субстанция — информацией». Далеко не все специалисты соглашались с таким определением, но оно, по крайней мере, дает общее представление о том, чем информация принципиально отличается от материи. В самом деле — если вы отдадите своему приятелю книгу (материальный объект), у вас этой книги не станет. А если вы расскажете ему, в каком магазине эта книга продается (то есть передадите информационный объект), то сведения об этом магазине появятся у вашего приятеля, но и у вас они тоже сохранятся.

- **легко изменяются** — если разработчик захочет исправить обнаруженную на сайте ошибку или внести дополнения, — он легко сможет это сделать, и все пользователи сайта тут же увидят изменения. А можно ли поступить так же, если хочется внести изменения в уже вышедшую книгу?

С другой стороны, сайтам присущи и недостатки информационных объектов.

纸 *Дополнительно*

Чтобы прочесть книгу или газету, вам не требуется ничего, кроме самой книги или газеты. А чтобы получить доступ к информационному объекту (телепрограмме, радиопередаче, сайту), нужны как минимум три технических устройства:

- **источник** — телепрограммы и радиопередачи транслируются из специально оборудованной студии. Аналогично, для сайта нужен веб-сервер, который обеспечивает его функционирование;
- **приемник** — чтобы просмотреть телепрограмму или прослушать радиопередачу, нужно включить телевизор или радиоприемник. Аналогично, чтобы получить доступ к сайту, нужно устройство, предназначенное для просмотра сайтов (например, персональный компьютер с установленным на нем браузером);
- **линия связи** — наконец, чтобы информационный объект попал от источника к приемнику, между ними должна быть проведена линия связи. Теле- и радиосигналы распространяются по сложной системе кабелей, передаются по воздуху, проходят через ретрансляционные узлы и другие устройства. Аналогично, данные от веб-сервера следуют к приемнику через разнообразные коммуникации — в том числе через те же, которые используются для теле- и радиовещания, телефонной связи и других сигналов.

Еще один недостаток — изменчивость и непостоянство, характерные для Интернета. Сайты появляются и исчезают, меняют свои адреса, содержание и интерфейс. Пользователь не может быть уверен в том, что завтра он посетит свой любимый сайт так же, как делал это вчера и сегодня.

В Сети полно «мертвых» ссылок, ведущих на давно исчезнувшие или изменившиеся сайты. Но особенно нелепо выглядят ссылки на сетевые ресурсы в печатных изданиях, тем более в тех, которые рассчитаны на долгую жизнь: в энциклопедиях, учебниках, словарях. «Мертвую» ссылку в энциклопедии нельзя исправить, и она будет сбивать с толку не одно поколение читателей.

1.2.2. Традиционное — новое

Тогда телевидение считалось новинкой; так что, даже если актер пукнет в кадре, все начинали говорить, что это свежая концепция и что ничего похожего до сих пор на телевидении не было.

Дж. Бернс

Печатные издания. История развития — несколько столетий, а если учитывать изготовление документов на бумаге до изобретения книгопечатания, — несколько тысячелетий.

Существуют общепринятые нормы, правила и традиции, подкрепленные многовековым опытом. Большинство нововведений опираются на эти правила и традиции. Разработаны государственные и международные стандарты, сложилась общепринятая терминология.

Сайты. История развития — менее 15 лет.

Существуют некоторые общепринятые стандарты (например, протокол HTTP, схема именования ресурсов URL), но в целом дела со стандартизацией обстоят плохо. Браузеры по-разному отображают один и тот же код HTML и CSS, по-разному обрабатывают сценарии JavaScript и другие технологии. Разработчикам сайтов приходится прилагать усилия, чтобы сайты были совместимы с различными браузерами.

纸 *Дополнительно*

Нет единообразия и в подходе к пользовательскому интерфейсу сайтов. Здесь уместно провести параллель с графическим интерфейсом традиционного программного обеспечения. Какую бы программу мы ни взяли — хоть текстовый редактор, хоть почтовый клиент, хоть файловый менеджер, хоть браузер, — всегда сверху мы увидим привычную строку меню с традиционными пунктами **Файл**, **Правка**, **Вид** и т. д., всегда в меню **Файл** можно найти команды **Открыть** и **Сохранить как**, а в меню **Правка** — **Вырезать** и **Копировать**. Под строкой меню всегда находится панель инструментов, в нижней части окна — строка состояния, справа — полоса прокрутки, в правом верхнем углу — кнопки **Заккрыть** и **Свернуть**. Все стандартные элементы интерфейса в разных программах имеют один и тот же цвет и размер, команды меню и другие надписи отображаются одним и тем же шрифтом. Поэтому, если пользователь работал