

Могилев А. В.

Листрова Л. В.

**ТЕХНОЛОГИИ ОБРАБОТКИ
ТЕКСТОВОЙ ИНФОРМАЦИИ**

**ТЕХНОЛОГИИ ОБРАБОТКИ
ГРАФИЧЕСКОЙ И МУЛЬТИМЕДИЙНОЙ
ИНФОРМАЦИИ**

Санкт-Петербург

«БХВ-Петербург»

2010

УДК 681.3.06(075.3)
ББК 32.973.26-018.2я721
М74

Могилев, А. В.

М74 Технологии обработки текстовой информации. Технологии обработки графической и мультимедийной информации / А. В. Могилев, Л. В. Листрова. — СПб.: БХВ-Петербург, 2010. — 304 с.: ил. — (ИиИКТ, Профильная школа)

ISBN 978-5-9775-0468-3

Книга является частью комплекта учебников по курсу информатики и информационно-коммуникационных технологий (ИКТ) в старших классах общеобразовательной школы на профильном уровне. Она охватывает 7-й и 8-й модули курса и является продолжением пособий "Информация и информационные процессы. Социальная информатика", "Средства информатизации. Телекоммуникационные технологии", "Методы программирования. Компьютерные вычисления".

В книге рассмотрены: принципы и технологии автоматизированной обработки текстовой информации, текстовые редакторы и их возможности, издательские технологии и терминология, гипертекстовое представление информации. Описаны системы автоматического анализа и синтеза текста (лингвистические, распознавания речи и др.), системы автоматизированного перевода. Раскрыты понятия мультимедиа и медиакультуры, компьютерной графики и дизайна, описана работа с графической, звуковой и видеoinформацией, рассмотрены мультимедийные презентационные технологии.

По каждой рассматриваемой теме предлагаются контрольные вопросы, темы для рефератов и докладов, вопросы для обсуждения, задачи и упражнения, лабораторные работы.

*Для учащихся старших классов физико-математического,
информационно-технологического и других профилей*

УДК 681.3.06(075.3)
ББК 32.973.26-018.2я721

Группа подготовки издания:

Главный редактор	<i>Екатерина Кондукова</i>
Зам. главного редактора	<i>Людмила Еремеевская</i>
Зав. редакцией	<i>Григорий Добин</i>
Редактор	<i>Екатерина Капалыгина</i>
Компьютерная верстка	<i>Натальи Караваевой</i>
Корректор	<i>Виктория Пиотровская</i>
Дизайн серии	<i>Инны Тачиной</i>
Оформление обложки	<i>Елены Беляевой</i>
Зав. производством	<i>Николай Тверских</i>

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 31.05.10.

Формат 70×100^{1/16}. Печать офсетная. Усл. печ. л. 24,51.

Тираж 1000 экз. Заказ №

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

Санитарно-эпидемиологическое заключение на продукцию № 77.99.60.953.Д.005770.05.09 от 26.05.2009 г. выдано Федеральной службой по надзору в сфере защиты прав потребителей и благополучия человека.

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

Оглавление

Предисловие.....	1
МОДУЛЬ 7. ТЕХНОЛОГИИ ОБРАБОТКИ ТЕКСТОВОЙ ИНФОРМАЦИИ	5
РАЗДЕЛ 7.1. ПРИНЦИПЫ И ТЕХНОЛОГИИ АВТОМАТИЗИРОВАННОЙ ОБРАБОТКИ ТЕКСТОВОЙ ИНФОРМАЦИИ	7
Учебный материал	9
Сканирование и распознавание текста.....	12
Структурирование текста	13
Операции смысловой обработки текста.....	14
Обработка "специальных" текстов	15
Контрольные вопросы	17
Темы для рефератов.....	17
Вопрос для обсуждения.....	17
Задачи и упражнения	18
Лабораторные работы.....	18
РАЗДЕЛ 7.2. ТЕКСТОВЫЕ ПРОЦЕССОРЫ.....	19
7.2.1. Назначение и основные функции текстовых процессоров	21
Учебный материал	21
Контрольные вопросы	25
Темы для рефератов.....	26
Вопросы для обсуждения.....	26
Задачи и упражнения	26
Лабораторные работы.....	27
7.2.2. Создание текстовых документов	28
Учебный материал	28
Набор текста, вставка символов	29

Вставка и редактирование графических объектов.....	32
Создание маркированных и нумерованных списков.....	34
Создание и оформление таблиц.....	36
Заполнение колонтитулов.....	40
Использование Microsoft Equation Editor.....	41
Специальная вставка объектов в текстовый документ.....	42
Сохранение документов в различных форматах.....	44
Создание форм.....	44
Слияние документов.....	45
Контрольные вопросы.....	47
Темы для рефератов.....	48
Вопросы для обсуждения.....	49
Задачи и упражнения.....	49
Лабораторные работы.....	50
7.2.3. Приемы редактирования и рецензирования текстовых документов.....	51
Учебный материал.....	51
Перемещение по документу.....	52
Выделение текстовых фрагментов.....	55
Редактирование текстового документа.....	57
Перенос слов.....	62
Рецензирование документов.....	63
Контрольные вопросы.....	66
Темы для рефератов.....	67
Вопросы для обсуждения.....	67
Задачи и упражнения.....	67
Лабораторные работы.....	68
7.2.4. Форматирование и подготовка документа к печати.....	69
Учебный материал.....	69
Основные операции по форматированию текста.....	69
Использование стилевого оформления документа.....	76
Создание оглавлений.....	78
Обработка текстовых документов с помощью макросов.....	79
Вывод документа на печать.....	81
Контрольные вопросы.....	84
Темы для рефератов.....	85
Вопросы для обсуждения.....	85
Задачи и упражнения.....	85
Лабораторные работы.....	86

РАЗДЕЛ 7.3. ИЗДАТЕЛЬСКИЕ СИСТЕМЫ	89
7.3.1. Издательские технологии и программное обеспечение	91
Учебный материал	91
Microsoft Word	95
Microsoft Publisher	96
Adobe PageMaker	96
TeX	97
QuarkXPress	97
Другие программы	98
Контрольные вопросы	98
Темы для рефератов	99
Вопросы для обсуждения	99
Задачи и упражнения	100
Лабораторные работы	100
7.3.2. Шрифты	101
Учебный материал	101
Контрольные вопросы	107
Темы для рефератов	108
Вопросы для обсуждения	108
Задачи и упражнения	109
Лабораторные работы	109
7.3.3. Верстка и иллюстрирование текста	111
Учебный материал	111
Основные правила верстки	113
Правила набора текста	113
Приемы верстки иллюстраций	116
Включение в издание информационной графики	118
Основные правила дизайна	119
Шрифтовая политика	120
Контрольные вопросы	123
Темы для рефератов	124
Вопросы для обсуждения	124
Задачи и упражнения	124
Лабораторные работы	125
7.3.4. Создание печатных материалов (публикаций) в издательской системе Microsoft Publisher	126
Учебный материал	126
Настройка параметров бюллетеня	127

Контрольные вопросы	132
Темы для рефератов.....	132
Вопросы для обсуждения.....	133
Задачи и упражнения	133
Лабораторные работы.....	133

7.3.5. Технология разработки документов в издательской системе PageMaker

Учебный материал	134
Окно программы, основные команды	135
Создание страниц-шаблонов.....	136
Верстка страниц	138
Контрольные вопросы	140
Темы для рефератов.....	140
Вопросы для обсуждения.....	140
Задачи и упражнения	141
Лабораторные работы.....	142

РАЗДЕЛ 7.4. ГИПЕРТЕКСТ И ЯЗЫКИ ГИПЕРТЕКСТОВОЙ РАЗМЕТКИ

Учебный материал	145
Метаязык SGML	146
Язык гипертекстовой разметки HTML	149
Язык разметки XML.....	161
Контрольные вопросы	170
Темы для рефератов и докладов.....	171
Вопросы для обсуждения.....	171
Задачи и упражнения	171
Лабораторные работы.....	171

РАЗДЕЛ 7.5. ТЕХНОЛОГИИ ОБРАБОТКИ ТЕКСТА, ИСПОЛЬЗУЮЩИЕ МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Учебный материал	175
Автоматическая проверка правописания	177
Голосовой ввод текста	177
Машинный перевод на иностранные языки	180
Контрольные вопросы	185
Темы для рефератов и докладов.....	185
Вопросы для обсуждения.....	186
Задачи и упражнения	186
Лабораторные работы.....	186

**МОДУЛЬ 8. ТЕХНОЛОГИИ ОБРАБОТКИ ГРАФИЧЕСКОЙ
И МУЛЬТИМЕДИЙНОЙ ИНФОРМАЦИИ 187****РАЗДЕЛ 8.1. ПОНЯТИЕ МУЛЬТИМЕДИА..... 189**

Учебный материал	191
Контрольные вопросы	193
Темы для рефератов и докладов.....	194
Вопросы для обсуждения.....	194
Задачи и упражнения	194
Лабораторные работы.....	195

**РАЗДЕЛ 8.2. ПРИНЦИПЫ И ТЕХНОЛОГИИ ОБРАБОТКИ
ГРАФИЧЕСКОЙ ИНФОРМАЦИИ 197**

CorelDraw.....	197
Учебный материал	199
Контрольные вопросы	207
Темы для рефератов и докладов.....	208
Вопросы для обсуждения.....	208
Задачи и упражнения	208
Лабораторные работы.....	209

8.2.1. Графический редактор Paint 211

Учебный материал	211
Контрольные вопросы	221
Темы для рефератов.....	221
Вопросы для обсуждения.....	221
Задачи и упражнения	222
Практическая работа	222

**8.2.2. Профессиональные графические редакторы Adobe Photoshop
и CorelDraw 223**

Учебный материал	223
Контрольные вопросы	233
Темы для рефератов.....	233
Вопросы для обсуждения.....	233
Задачи и упражнения	234
Практическая работа	234

РАЗДЕЛ 8.3. ПРИНЦИПЫ И ТЕХНОЛОГИИ ОБРАБОТКИ ЗВУКОВОЙ ИНФОРМАЦИИ	235
Учебный материал	237
Контрольные вопросы	245
Темы для рефератов.....	246
Вопросы для обсуждения.....	246
Задачи и упражнения	246
Практическая работа	247
РАЗДЕЛ 8.4. ПРИНЦИПЫ КОМПЬЮТЕРНЫХ ВИДЕОТЕХНОЛОГИЙ И АНИМАЦИИ	249
Учебный материал	251
Оборудование для цифровых видеотехнологий.....	251
Форматы видеофайлов.....	254
Программные средства для нелинейного видеомонтажа.....	259
Мультимедийная инструментальная среда Macromedia Flash фирмы Macromedia.....	262
Контрольные вопросы	267
Темы для рефератов.....	267
Вопросы для обсуждения.....	268
Задачи и упражнения	268
Практические работы	268
РАЗДЕЛ 8.5. МУЛЬТИМЕДИЙНЫЕ ПРЕЗЕНТАЦИОННЫЕ ТЕХНОЛОГИИ, ИХ ИСПОЛЬЗОВАНИЕ В УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ	269
Учебный материал	271
Создание структуры презентации.....	272
Основные приемы работы с программой Microsoft PowerPoint.....	272
Контрольные вопросы	286
Темы для рефератов.....	286
Вопросы для обсуждения.....	286
Задачи и упражнения	287
Лабораторные работы.....	288
Предметный указатель	289

Предисловие

Несмотря на значительные усилия, направленные на становление профильного образования, изменения в российской школе еще далеки от ожиданий. Одна из причин связана с недостаточностью учебно-методического обеспечения обучения на профильном уровне во многих образовательных областях, к которым относятся также информатика и информационные технологии.

Данный цикл пособий призван хотя бы отчасти снизить остроту этой проблемы и ликвидировать пробел в учебной литературе, предназначенной для обучения старшеклассников по профилям, включающим курс информатики и информационно-коммуникационных технологий профильного уровня: информационно-технологическом, физико-математическом и др.

Мы предлагаем модульный вариант как самого курса информатики и информационно-коммуникационных технологий, так и его учебного обеспечения. Значительный объем всего профильного курса (280 часов), разнородность его содержательных компонентов (как с точки зрения соотношения теоретических и практических вопросов, так и сложности и трудоемкости деятельности учащихся), быстрое и неоднородное их развитие диктуют именно модульное построение данного курса, позволяющее выбирать, использовать и развивать модули учебного пособия независимо друг от друга. Кроме того, модульная структура курса обеспечивает большую гибкость и эффективность в достижении качества обучения, поскольку оно гарантируется качеством освоения каждого из модулей в отдельности.

Согласно авторской учебной программе данный профильный учебный курс состоит из следующих 10 модулей:

1. Информация и информационные процессы.
2. Социальная информатика.
3. Средства информатизации.
4. Телекоммуникационные технологии.

5. Методы программирования.
6. Компьютерные вычисления.
7. Технологии обработки текстовой информации.
8. Технологии обработки графической и мультимедийной информации.
9. Технологии поиска и хранения информации.
10. Технологии автоматизации управления.

Настоящее пособие охватывает 7-й и 8-й модули курса и является продолжением пособий "Методы программирования. Компьютерные вычисления" тех же авторов.

Данная книга посвящена информационным технологиям обработки текста, звука, графики и видео. Эти технологии являются целью развития информатики. Их появление и развитие обусловили появление широкого круга пользователей этих технологий различных уровней от начинающих любителей до высококвалифицированных профессионалов, дали толчок широкому применению компьютера в различных сферах нашей жизни, а также огромной распространенности самих компьютеров и такому глубокому усовершенствованию, которое они прошли за исторически недолгое время с момента их появления.







Мы рассмотрим технологии, реализованные с помощью различных программных средств на основе IBM-совместимых персональных компьютеров и операционной системы MS Windows.

Владение технологиями обработки текста, звука, графики и видео относится к содержанию подготовки пользователя (или продвинутого пользователя) ПК.

Описание каждой технологии сопровождается обзором реализующих ее программных средств. По этим технологиям и программным средствам приводятся краткие сведения и примеры, чтобы сложилась общая картина и стало возможным дальнейшее самостоятельное углубленное изучение этих технологий на основе специальной литературы, интернет-источников и электронных обучающих средств.

Пособия содержат учебный материал и различного рода задания и вопросы: тексты для чтения и изучения, контрольные вопросы, темы рефератов, вопросы для обдумывания и обсуждения, задания и упражнения, а также описание лабораторных работ. Это позволяет организовать занятия различного типа и реализовывать различные методы обучения: от лекций и семинарских занятий до учебных проектов.

Для того чтобы облегчить ориентацию в пособии и придать ему структуру, мы используем следующие обозначения:

-  — Учебный материал;
-  — Контрольные вопросы;
-  — Темы для рефератов и докладов;
-  — Вопросы для обсуждения;
-  — Задачи и упражнения;
-  — Лабораторные или Практические работы.

МОДУЛЬ 7

Технологии обработки текстовой информации

- 7.1. Принципы и технологии автоматизированной обработки текстовой информации**
- 7.2. Текстовые процессоры**
- 7.3. Издательские системы**
- 7.4. Гипертекст и языки гипертекстовой разметки**
- 7.5. Технологии обработки текста, использующие методы искусственного интеллекта**

РАЗДЕЛ 7.1

**Принципы и технологии
автоматизированной обработки
текстовой информации**



Учебный материал

Широкое распространение персональных компьютеров позволило совершить поистине революционный переворот в автоматизации офисного делопроизводства в 80-е годы прошлого века. Персональные компьютеры в первую очередь вытеснили из обращения печатающие машинки, составляющие основу технических средств подготовки печатных документов. Преимущества компьютера, оснащенного специальным текстовым процессором (редактором), перед печатающей машинкой были явными и заключались в том, что обеспечивали значительное повышение удобства, производительности выполнения работ и, самое главное, повышение качества получаемых при этом документов. Разделение во времени этапов подготовки документа, таких как ввод, редактирование, оформление, подготовка к печати и собственно сама печать сделали процесс создания документа более простым и технологичным.

В настоящее время текстовые процессоры предоставляют пользователю широкие возможности по автоматизации процесса подготовки документов. Это и функции редактирования, допускающие возможность любого изменения, вставки, замены, копирования и перемещения фрагментов в рамках одного документа и между различными документами, контекстного поиска, функции форматирования символов, абзацев, страниц, разделов документа, верстки, проверки грамматики и орфографии, использования наряду с простыми текстовыми элементами списков, таблиц, рисунков, графиков и диаграмм.

Значительное сокращение времени подготовки документов обеспечивают такие средства автоматизации набора текста, как *автотекст* и *автозамена*, использование форм, шаблонов и мастеров типовых документов. Процедуру ввода данных значительно упрощают сканеры и голосовые устройства. Существующие системы распознавания текстов, принимаемых со сканера,

включают функцию экспорта документа в текстовые редакторы. Наличие внешней памяти компьютера обеспечивает удобное длительное хранение подготовленных ранее документов, быстрый доступ к ним в любое время. Широкий спектр печатающих устройств в сочетании с функциями подготовки документа к печати, предварительного просмотра, обеспечивает получение высококачественных черно-белых и цветных копий на бумаге и прозрачной пленке.

Тенденции совершенствования автоматизированных систем обработки текстовой информации в настоящее время направлены на улучшение коммуникационных возможностей текстовых процессоров. Так, при работе в локальных и глобальных компьютерных сетях имеется возможность обмениваться документами с удаленными пользователями, отправлять документы по электронной почте непосредственно из текстового процессора, создавать документы в формате Web-страниц.

Современные текстовые процессоры обладают также мощными средствами создания макрокоманд, которые реализуются на развитом языке сценариев, позволяющем решать довольно сложные задачи преобразования и оформления публикаций.

Устройства автоматизированной обработки текстовой информации принципиально отличаются от устройств, используемых для обработки данных. Это обусловлено тем, что обработка текста, выраженного словами и предложениями, не может осуществляться только в пакетном режиме (без участия человека). Составление и редактирование текста может проводиться только человеком, взаимодействующим с компьютером. При подготовке текстовых документов на компьютере используются три основные группы операций: *ввод, редактирование и форматирование*.

Операции ввода позволяют перевести исходный текст из его внешней формы в электронный вид, то есть в файл, хранящийся на компьютерном носителе информации. Под вводом не обязательно понимается машинописный набор с помощью клавиатуры. Ввод текста можно осуществить также и путем сканирования бумажного или пленочного оригинала, а затем с помощью программы распознавания образов перевести документ из формата графического изображения в текстовый формат.

В настоящее время устройства автоматизированной обработки текстовой информации на крупных предприятиях широко используются не только операторами, но и менеджерами всех уровней, которые на изображенном на экране документе могут делать свои замечания, ставить резолюции, что значительно упрощает процедуру согласования их действий и ускоряет процесс подготовки управленческих решений. Использование диктофонной техники, установленной в центре автоматизированной обработки текстовой информации, облегчает менеджеру подготовку любых документов, поскольку, подключив

шись к центру по телефонной или селекторной связи и продиктовав необходимый материал, он в кратчайшие сроки может получить его распечатку или перевод на иностранный язык.

Операции редактирования (правки) позволяют изменить уже существующий электронный документ путем добавления или удаления его фрагментов, перестановки частей документа, слияния нескольких файлов в один или, наоборот, разбиения единого документа на несколько более мелких. Ввод и редактирование при работе над текстом часто выполняют параллельно.

При вводе и редактировании формируется содержание текстового документа. Оформление документа задают *операциями форматирования*. Команды форматирования позволяют точно определить, как будет выглядеть текст на экране монитора или на бумаге после его распечатки на принтере.

Все электронные текстовые документы требуют ввода и, обычно, редактирования, операция форматирования документа не всегда является обязательной. Более того, форматирование текстового документа в некоторых случаях вредит делу, поскольку информация о форматировании заносится в текст в виде невидимых кодов, которые могут "мешать" определенным программам работать с текстами. Так, например, программа Блокнот не способна отразить на экране текстовый файл, созданный в редакторе WordPad, хотя оба приложения принадлежат к одной группе стандартных программ и оба предназначены для работы с текстами.

Для автоматизированной обработки текстовой информации имеются два типа программ, одни из которых используются только для ввода и редактирования текста (*текстовые редакторы*), а другие позволяют также его форматировать (*текстовые процессоры*).

Все текстовые редакторы сохраняют в файле "чистый" текст и, благодаря этому, совместимы друг с другом. Различные текстовые процессоры информацию о форматировании записывают в файл по-разному и поэтому несовместимы друг с другом. Однако во многих текстовых процессорах есть возможность преобразования текста из одного формата в другой.

При выборе текстового процессора для работы необходимо учитывать такие факторы, как сложность создаваемых документов, характер размещаемых в документах материалов (например, обычный текст, таблицы, формулы и т. п.), масштаб (объемы) текстов, требования к качеству распечатки документа на бумаге и др. Наиболее известные текстовые процессоры "по специализации" можно условно разделить на три группы:

- процессоры общего назначения (Microsoft Word, OpenOffice Writer);
- редакторы научных документов (ChiWriter, TeX);
- редакторы исходных текстов программ (Multi-Edit и встроенные редакторы систем программирования Basic, Pascal и т. д.).

Разумеется, с помощью MS Word можно подготовить и текст программы, а с помощью Multi-Edit — документ общего назначения. "Специализация" процессора заключается в том, что в нем добавлены (или оптимизированы) функции, необходимые для обработки документов определенного типа. Например, Multi-Edit позволяет выделять цветом смысловые сегменты исходных текстов программ, TeX удобен для набора математических выражений и т. д.

Особое место в группе пакетов, предназначенных для автоматизированной обработки текстовой информации, занимают так называемые издательские системы — Aldus PageMaker, Corel Ventura, QuarkXPress и др. Как правило, издательские системы используются только для подготовки набранного документа к тиражированию (верстка и макетирование издания). Для набора текста все же удобнее применять текстовые процессоры, а для создания и редактирования иллюстраций — графические системы (CorelDRAW, Adobe Photoshop и т. д.).

Сканирование и распознавание текста

Одной из наиболее распространенных программ, предназначенных для сканирования и распознавания текста, является программа FineReader. Простейшая из команд этой программы — **Сканировать и распознать** из меню **Процесс**. По этой команде будет отсканирована и тут же распознана одна страница текста. После завершения процесса распознавания распознанный текст появится в окне **Текст**, после чего его можно читать и править, как в обычном редакторе текстов.

При сканировании целой книги удобно сначала отсканировать все ее страницы, а потом их распознать. Для этого нужно воспользоваться командой **Сканировать несколько страниц** меню **Файл**, а затем — командой **Распознать все** меню **Процесс**. Перед сканированием разворота книги в настройках сканера необходимо установить признак альбомной ориентации изображения, а на вкладке **Сканирование** диалога **Опции** — **Делить Разворот Книги**. Это приведет к тому, что после сканирования FineReader разделит два листа разворота на две разных страницы и будет распознавать их отдельно. Такой метод улучшит качество распознавания и уменьшит количество ошибок. При сканировании отдельных страниц эту опцию можно убрать.

Текст "нормального" размера и хорошего качества печати можно сканировать при **Ручном подборе яркости** 50 %. В **Режиме сканирования картинок** нужно установить **Черно-белое изображение** при **Разрешении** 300 (это самый быстрый режим сканирования). Для мелких шрифтов разрешение можно увеличить до 400, 500 или 600. Если качество печати плохое или книга совсем плохо разворачивается, необходимо перейти к **Автоматическому Подбору Яркости** и в **Режиме Сканирования Картинок** установить **Серый**.

Это замедлит сканирование, но улучшит качество распознавания. Все описанные параметры устанавливаются в окне диалога **Настройки Сканера**.

Если на листе встречаются серые пятна, или из-за тонкой бумаги и жирного черного текста проступают буквы и рисунки с обратной стороны страницы, увеличьте яркость сканирования, например, до 60 или до 70. Это приведет к исчезновению пятен и проступающих с обратной стороны страницы букв и рисунков.

Если буквы слишком бледные, то уменьшите яркость сканирования, например, до 40. Это приведет к тому, что буквы станут более четкими и жирными, и их легче будет распознавать.

Если текст на листе расположен в несколько колонок, то на вкладке **Сканирование** диалогового окна **Опции** переключатель **Тип Страницы** необходимо установить в положение **Авто**. А если это обычный книжный текст, то лучше поставить переключатель **Тип Текста** в значение **Одна Колонка**, что ускорит процесс распознавания.

Структурирование текста

При создании текстовых документов большое значение имеет пространственное расположение текста, способствующее более благоприятному его восприятию. Единицей пространственного размещения текстовой информации служит *абзац* (выделенный по смыслу участок текста). Для оформления абзаца используют несколько параметров:

- *Выравнивание (выключка)* — правило расположения букв в строке абзаца. Видов выравнивания четыре: по левому краю, центральное, по правому краю и по ширине полосы набора.
- *Отступы от краев полосы набора*.
- *Абзацный отступ (красная строка)* — положение первой строки абзаца.
- *Интервалы*. Различают межстрочное расстояние, которое задается множителем размера шрифта (одинарный, полуторный, двойной интервал), и промежутки до и после абзаца.
- *Буквица* — крупная выступающая первая буква абзаца. Часто задается не просто более крупным размером буквы, но и буквой другого рисунка.

Абзацы размещаются в рамках полосы — выделенного участка страницы, как правило, прямоугольной формы, в котором размещаются текст и иллюстрации. На листе может быть либо одно такое место (одна колонка), либо несколько (в этом случае говорят о многоколоночном тексте). Как правило, текстовые процессоры не дают появляться висячим строкам — отдельным строкам абзацев в начале или конце полосы.

Важным элементом оформления текста на странице являются поля — пробелы вдоль края страницы и интервалы между колонками. Для удобочитаемости такие пробелы должны быть обязательно.

Как правило, в достаточно большом тексте выделяется несколько смысловых блоков (разделов) и видов содержательного текста — обычный текст, примечания, ссылки и т. п. Для оформления таких типовых элементов создаются *стили* — определенные наборы параметров оформления шрифта и абзацев. Применение стилей позволяет ускорить набор, автоматизировать оформление (например, автоматически создавать оглавления) и изменять внешний вид различных элементов, не разыскивая их по всему тексту. Практически все современные текстовые процессоры опираются на стили, даже если пользователь не использует их.

Для решения некоторых типовых задач оформления текстов существующие текстовые процессоры предусматривают возможность создания *списков и таблиц*.

Под *списком* понимают набор визуально выделенных элементов перечисления, которые выделяют либо с помощью символа-маркера (маркированные списки), либо номером (в упорядоченных списках). При оформлении списка чаще всего также предусматривают форматирование абзацев так, чтобы они не выступали за маркер. Автоматизация оформления списков позволяет автоматически менять маркеры и добавлять новые элементы списков.

Применение таблиц позволяет представлять информацию в виде двумерных структур, каждая ячейка которых может быть оформлена практически в виде листа в миниатюре.

Операции смысловой обработки текста

Текстовое представление информации — одна из наиболее удобных форм, позволяющая автоматизировать процессы ее смысловой обработки. К числу наиболее распространенных автоматизированных операций смысловой обработки текста относятся поиск, расшифровка или уточнение значений слов, реферирование и перевод.

Задача поиска необходимой информации чаще всего формулируется как поиск фрагментов, содержащих некоторые понятия, в достаточно большом массиве. Большое значение этот вид автоматической обработки получил с ростом популярности сети Интернет. Существует несколько подходов к организации такого поиска.

Суть первого подхода заключается в поиске фрагмента текста, соответствующего заданному образцу. Наиболее популярная форма задания такого образца — так называемые *регулярные выражения*. По сути, это описание

фрагмента текста, удовлетворяющего некоторым условиям, по тем частям, которые в нем содержатся, и их порядку. Таким способом в большом текстовом массиве можно находить упоминания тех или иных слов, адреса, номера телефонов и т. п. шаблонные элементы. Достоинством этого подхода является возможность его применения без предварительной обработки текста. Например, сразу при посимвольном получении текста.

Второй подход предусматривает предварительное создание специального вида базы для ускорения поиска — *индекса*. Такой способ применяется для ускорения поиска, если некоторые типовые поисковые запросы повторяются часто, и нет возможности формировать/хранить весь массив текста. Например, при организации поисковой машины в среде Интернет.

Для решения задачи расшифровки или уточнения значений слов применяют *словари* — базы информационных фрагментов, связанных с некоторыми ключевыми словами или словосочетаниями. Примером таких баз могут служить словари различных языков: англо-русский, русско-английский, толковый и другие виды словарей. Одним из самых распространенных применений словарей является проверка правописания слов при их наборе.

Особым видом словарей являются *тезаурусы* — словари, в которых слова связываются на основе каких-либо лексических отношений. Например: слова являются синонимами (смысловыми аналогами), антонимами (противоположны по смыслу) и т. п. Этот вид словарей также важен для систем, моделирующих отдельные аспекты человеческого мышления.

Используя закономерности естественного языка и описанные ранее средства выполнения некоторых операций и выявления зависимостей, с помощью компьютера автоматизируют (хотя и не полностью) некоторые операции по смысловому преобразованию текста. Современные системы обработки позволяют создавать краткие обзоры текстов (рефераты) или готовить перевод с одного естественного языка на другой.

Следует заметить, что точного решения задачи смысловой обработки текста не имеют, поскольку зачастую при переводе текста трудно подобрать адекватное слово с учетом грамматических особенностей, культурных традиций и т. п. Тем не менее, с применением специализированных по областям знания словарей системы автоматизированного перевода создают подстрочник, который дает представление о смысле текста и в дальнейшем может помочь в переводе документа.

Обработка "специальных" текстов

Под "*специальными*" текстами подразумевают тексты, содержащие математические, химические или другие формулы, сложные схемы и специфические обозначения, используемые в научных, учебных и технических публикациях

и документах. Для создания таких фрагментов стандартные средства представления и подготовки текста плохо приспособлены.

Существует множество специальных программных средств, предназначенных для подготовки специальных текстов. Наиболее популярным способом интеграции элементов-формул в документы является технология OLE. Технология предусматривает, что в документе выделяется место для размещения объекта, а обработка его ведется с помощью внешней программы, выступающей как OLE-сервер.

Такой способ позволяет интегрировать в одном документе разные объекты, но для корректной обработки и печати требует наличия соответствующих программ, а для редактирования — большое количество системных ресурсов.

При подготовке научных, технических и учебных текстов часто используется свободно доступная система подготовки публикаций TeX (от гр. *teRcnh* — "искусство", "мастерство"). При использовании этой системы документ с формулами описывается на специальном языке разметки в виде текстового файла, который и обрабатывается системой. Результатом становится специальный файл (*dvi*, *device independent* — "независимый от устройства"), который может быть просмотрен, напечатан или преобразован в другой формат с помощью специальных программ из комплекта.

Для соблюдения стандартов и упрощения набора систему комплектуют набором шаблонов и указаний о формировании страниц. Примерами таких шаблонов являются комплекты LaTeX, MikiTeX, AMSTeX.

Файл с материалом для этой системы набора может быть подготовлен с помощью обычного текстового редактора и передан на любую другую платформу. Тексты, подготовленные с помощью этой системы, соответствуют строгим стандартам оформления формул и научных текстов. Многие системы визуального набора позволяют сохранять описания формул в стандарте одного из комплектов TeX. Так, "облегченная" версия специализированного редактора MathType, входящая в пакет MS Office под названием Equation, позволяет набирать математические формулы при подготовке документов в текстовом процессоре Word.

В современных условиях все большее значение приобретает отображение документа с помощью браузеров web-страниц с минимальным количеством дополнительных средств. Для решения этой задачи в общем стандарте XML предусмотрен язык специальной разметки MathML. Формулы на этом языке описываются и отображаются в документах с помощью дополнительных модулей к программам-обозревателям.



Контрольные вопросы

1. Какие возможности по автоматизации процесса подготовки документов предоставляют пользователю современные текстовые процессоры?
2. Охарактеризуйте основные группы операций, используемых при подготовке текстовых документов на компьютере.
3. Приведите примеры текстовых процессоров из различных групп "специализации".
4. Какие параметры настройки сканера (и в каких случаях) позволяют улучшить скорость сканирования и качество распознавания текста?
5. Какие параметры и способы структурирования текста способствуют более благоприятному его восприятию?
6. Охарактеризуйте основные операции смысловой обработки текста.
7. Приведите примеры программных средств и технологий обработки "специальных" текстов.



Темы для рефератов

1. Системы локального поиска.
2. Системы и утилиты автоматизированной обработки текста.
3. Основные возможности программы сканирования и распознавания текста FineReader.
4. Алгоритмы и схемы сканобработки.
5. Системы оптического распознавания текстов.
6. Технологии оптического распознавания символов.
7. Текстовый процессор OpenOffice Writer.



Вопрос для обсуждения

Можно ли полностью автоматизировать процесс подготовки и обработки текстовых документов?



Задачи и упражнения

1. Запишите алгоритм сканирования, распознавания и обработки текстового документа.
2. Используя программу сканирования и распознавания текста (например, FineReader), отсканируйте страницу типографского текста и страницу текста, распечатанного на принтере. Конвертируйте полученные файлы в формат Microsoft Word для последующей обработки. Отредактируйте и отформатируйте документы в соответствии с исходными.



Лабораторные работы

1. Отсканируйте и распознайте два листа журнального текста, на которых для разных статей были применены различные фоны, имеются инвертирования, разбиения на блоки, повороты текста, а также таблицы. В полученных электронных документах проверьте орфографию. Таблицы сохраните в формате MS Excel, а текст в формате RTF.
2. Проведите эксперимент по сканированию и распознаванию рукописного текста.

РАЗДЕЛ 7.2

Текстовые процессоры

- 7.2.1. Назначение и основные функции текстовых процессоров**
- 7.2.2. Создание текстовых документов**
- 7.2.3. Приемы редактирования и рецензирования текстовых документов**
- 7.2.4. Форматирование и подготовка документа к печати**

7.2.1. Назначение и основные функции текстовых процессоров



Учебный материал

Класс прикладных программ, предназначенных для подготовки текстовых материалов на компьютере, называется *текстовыми редакторами*.

Поскольку текстовые материалы различаются сложностью набора и верстки (компоновки), существуют и развиваются различные программы обработки текстов. Эти программы можно классифицировать как по уровням возможностей по обработке текстов, так и по уровням требований, предъявляемых к составу и конфигурации компьютерной техники при их установке.

Текстовый редактор — программа для ввода и редактирования текстовых данных.

Текстовый процессор — программа для ввода, редактирования и форматирования текстовых данных.

Разработчики текстовых редакторов стараются предусмотреть в них предоставление пользователю всех необходимых операций и сервисных возможностей для эффективной обработки текстов. Выделим главные из них:

- набор текста в интерактивном режиме;
- редактирование текста;
- работа с фрагментами текста (копирование, перемещение, удаление и т. п.);
- форматирование текста (установка абзаца, перенос, выравнивание границ строки и т. п.);
- проверка правописания;
- поиск и замена фрагментов текста;
- работа с разными шрифтами;

- работа со спецсимволами (математические знаки, индексы и т. п.);
- работа с иллюстративным материалом (таблицы, схемы, формулы и пр.);
- работа с несколькими текстами одновременно посредством многооконного режима;
- файловая организация работы с текстами и взаимодействие с операционной системой;
- импорт/экспорт текстов из одного формата в другой, а также в другие прикладные системы.

Наиболее мощные из программ данного класса называются текстовыми процессорами. К примеру, программу NotePad (Блокнот) — стандартное Windows-приложение — уместно называть текстовым редактором, ее усовершенствованная версия WordPad занимает некоторое промежуточное положение между редактором и процессором, а мощная, широко популярная программа MS Word является текстовым процессором. Рассмотрим назначение и основные функции текстовых процессоров на примере этой программы.

Microsoft Word для Windows — это многофункциональный программный комплекс обработки текстов. Программа предназначена для выполнения работ по созданию документов, включающих разнообразные элементы (тексты, рисунки, графики, формулы, обычные или электронные таблицы, фрагменты баз данных, слайды и др.) и имеющих иерархическую организацию (разделы, главы, параграфы и т. п.).

Word позволяет автоматизировать работу пользователя для выполнения отдельных стандартных действий путем использования макрокоманд или программирования на языке *Visual Basic for Application*. Являясь компонентом Microsoft Office, Word может работать как изолированно, так и в составе пакета.

В первую очередь текстовый процессор MS Word предназначен для работы с текстом, а именно — со шрифтами. Возможно изменение их начертания, размера, цвета и стиля. Разнообразие самих шрифтов позволяет создавать документы самого разного типа: визитки, письма, брошюры, бланки, приглашения, объявления, вывески, книги и т. д. Доступна настройка всех параметров абзаца (отступ, межстрочный интервал, выравнивание).

MS Word позволяет создавать и видоизменять таблицы, а также проводить некоторые вычислительные и логические операции над числовыми данными. Процессор поддерживает работу с графикой. Для любого графического объекта можно настроить яркость, контрастность, размер, цвет, положение и другие параметры. Также есть возможность разработки высокохудожественных

надписей. Встроенный редактор формул позволяет создавать самые сложные математические и химические формулы.

Кроме всего перечисленного в Word имеется множество дополнительных функций, ускоряющих и облегчающих работу с текстами:

- расстановка переносов;
- проверка орфографии, грамматики и синтаксиса;
- поддержка нескольких словарей для проверки правильности написания;
- вставка в текст специальных символов;
- разработка своего стиля и использование его в дальнейшем для быстрого форматирования текста;
- создание границ вокруг текста и страниц;
- работа с макросами;
- поддержка многооконного режима;
- автоматическая нумерация страниц;
- настройка параметров страницы и поддержка нестандартных форматов бумаги;
- возможность предварительного просмотра документа перед печатью в удобном масштабе.

Главным объектом, создаваемым и редактируемым текстовым процессором, является *документ*. Современные текстовые процессоры позволяют создавать документы трех типов.

- Печатные документы* создаются и распечатываются на одном рабочем месте, и дальнейшее движение документа происходит только в бумажной форме.
- Электронные документы* существуют в виде файлов. Электронный документ, как правило, не является окончательным. Его можно дорабатывать, редактировать, форматировать, распечатывать или использовать для подготовки других документов.
- Web-документы*. Предполагается, что в этом качестве они останутся навсегда, и их преобразование в печатные документы не планируется. В Web-документах большую роль играет управление цветом. Для этой категории документов предполагается также наиболее широкий выбор средств форматирования и оформления.

Основными структурными элементами документа, подлежащими автоматическому форматированию, являются символ, абзац, страница, раздел.


Символ определяется видом, размером и начертанием шрифта.

Абзац — это часть текста, представляющая собой самую мелкую композиционную его единицу и графически обозначаемая в наборе пробелом слева в начальной строке (*абзацный отступ*). При наборе текста переход на следующую строку внутри абзаца происходит автоматически при полном заполнении текущей строки. По завершении абзаца нажатием клавиши <Enter> вводится специальный символ "конец абзаца" (¶).

Страница характеризуется размером бумажного листа и параметрами размещения текста: полями, отступами от колонтитулов, способами вертикального выравнивания, ориентации текста.

Раздел — это область документа, характеризующаяся определенным форматом печатной страницы; видом и содержанием колонтитулов; способом нумерации страниц; видом сносок в тексте и т. д. Количество разделов в документе не ограничено.

При вводе и редактировании текста работа осуществляется с документом в рабочем поле. Горизонтальная линейка форматирования над рабочим полем позволяет быстро изменять отступы абзацев, страничные поля, ширину колонок и табличных столбцов, а также устанавливать с помощью мыши позиции табуляции. Линейка устанавливается и убирается с помощью команды **Линейка** меню **Вид**. Единица измерения на линейке определяется установкой на вкладке **Общие** команды **Параметры** меню **Сервис**.

Горизонтальная и вертикальная полосы прокрутки используются для перемещения в документе по горизонтали и вертикали с помощью мыши. В нижней части вертикальной полосы прокрутки содержится специальная поисковая панель, позволяющая быстро переходить к конкретным объектам документа (страницам, рисункам, таблицам, сноскам и т. д.). Для ее открытия надо нажать кнопку **Выбор объекта**  движка вертикальной полосы прокрутки.

Для удобства работы с большими документами окно можно разбить на несколько частей (команда **Разделить** меню **Окно**) и просматривать в каждой из них отдельные фрагменты документа. Восстановление единого окна производится командой **Снять разделение** меню **Окно**.

В нижней части окна располагается строка состояния, содержащая различные сообщения, а также информацию о режимах работы с документом. Описание элементов и назначение индикаторов строки состояния представлены в табл. 7.1 и 7.2. Управление выводом строки состояния и полос прокрутки осуществляется на вкладке **Вид** команды **Параметры** меню **Сервис**.

Таблица 7.1

Элемент	Описание
Стр. 5	Номер текущей страницы
Разд. 1	Номер текущего раздела. Информация, отображаемая этим элементом, имеет смысл только для документов, разбитых на разделы
5/14	Номер текущей страницы (5) и общее количество страниц в документе (14)
На 7,8 см	Расстояние по вертикали между курсором и верхним краем страницы
Ст 10	Количество строк от начала страницы до положения курсора
Кол 48	Количество символов от левого поля до положения курсора, включая пробелы и символы табуляции

Таблица 7.2

Индикатор	Назначение
ЗАП	Активен режим записи макрокоманд
ИСПР	Активен режим маркировки исправлений. Маркеры исправления показывают, какие изменения были внесены в документ по сравнению с его последней версией
ВДЛ	Активна клавиша <F8> (Расширить выделение). Каждое нажатие клавиши <F8> будет расширять выделение фрагмента текста. Для сужения выделения используется комбинация клавиш <Shift>+<F8>
ЗАМ	Активен режим замены символов. В режиме вставки символов индикатор имеет серый цвет



Контрольные вопросы

1. Какие функциональные возможности отличают текстовый процессор от текстового редактора?
2. Перечислите функциональные возможности текстового процессора MS Word.
3. Какие виды документов можно создавать в текстовых процессорах? Каково назначение каждого из них?

4. Охарактеризуйте основные структурные элементы документа.
5. Для чего предназначены панели инструментов?
6. Как познакомиться с функциями, выполняемыми различными кнопками панелей инструментов?
7. Какая информация о документе и режимах его обработки отображается в строке состояния?



Темы для рефератов

1. История развития текстовых редакторов.
2. Системы обработки текстов в операционной системе MS DOS.
3. Текстовые процессоры семейства MS Word.
4. Системы обработки текстов в операционной системе Linux.
5. Свободно распространяемое ПО для обработки текстовой информации.



Вопросы для обсуждения

1. Для чего нужны простейшие текстовые редакторы (например, Блокнот)?
2. В чем основное преимущество текстового процессора MS Word по сравнению с программой WordPad?



Задачи и упражнения

1. Познакомьтесь с различными вариантами представления документа, имеющимися в меню **Вид**. В каких случаях удобно использовать каждый из них?
2. Настройте поля, абзацный отступ и позицию табуляции в имеющемся документе, используя горизонтальную и вертикальную линейки.