

## Кодирование и обработка текстовой информации

*Трапезникова Галина Альбертовна,  
учитель информатики и ИКТ ГБОУ школа №38  
Приморского района Санкт-Петербурга*

**Место темы в разделе и в курсе «Информатика и ИКТ»**  
Тема «Кодирование и обработка текстовой информации» изучается в базовом курсе информатики: в 8 классе по учебнику Н.Д. Угриновича в главе 1 «Информация и информационные процессы», в главе 2 «Кодирование и обработка текстовой информации» – в 9 классе по учебнику Н.Д. Угриновича. На изучение отводится 19 часов. Более глубоко в дальнейшем тема изучается в 10 классе. В 11 классе по данной теме материал прорабатывается во время изучения глав 1, 4 (создаются текстовые документы с информацией по темам параграфов) и главы 5 – итогового повторения.

### **Основные теоретические положения**

**Текстовая информация** – это информация, представленная в виде букв, знаков препинания и специальных символов некоторой **знаковой системы**. Буквы и другие значки принято называть символами. Набор их конечен. Текстовую информацию иногда называют символьной. Знаковая система содержит еще и правила выполнения операций над знаками (грамматика, синтаксис).

**Кодирование** – это процесс представления каждого символа в виде кода. **Код** – набор условных обозначений для представления информации. Количество знаков в коде называется **длиной кода**.

Естественные языки – это знаковые системы с исключениями из правил. Поэтому их нельзя использовать для кодирования информации с последующей обработкой на компьютере. Знаковые системы со строгими правилами называются формальными. Для компьютерного кодирования информации используется **формальная двоичная знаковая система**. Физическая природа знаков двоичного компьютерного кода – это электрические импульсы (наличие импульса или его отсутствие). При кодировании используется кодировочная таблица. Таблица устанавливает взаимно однозначное соответствие между знаками и их кодами.

Для русских букв существуют пять однобайтовых кодировочных таблиц:

1. Windows.
2. MS-DOS.
3. КОИ-8.
4. Mac.
5. ISO.

В них одному и тому же коду соответствуют разные символы. Текст, созданный в одной кодировке, не будет правильно отображаться в другой. С помощью указанных кодовых таблиц можно закодировать 256 символов. В современной кодировочной таблице Unicode, которая является международным стандартом, один символ кодируется 2 байтами. Таблица содержит 65536 символов, в том числе и символы русского алфавита.<sup>[1]</sup>

При нажатии на клавиши клавиатуры происходит процесс кодирования символов.

**Декодирование** – процесс обратный кодированию, т.е. код символа преобразуется в его изображение. Процесс декодирования информации осуществляется при выводе информации из оперативной памяти компьютера на экран монитора, например, или на листинг с помощью принтера.

**Символ** – наименьшая единица текстовой информации.<sup>[4]</sup>

**Слово** – это набор символов, ограниченный пробелами или знаками препинания.

**Строка** (не предложение) – это произвольная последовательность символов внутри абзаца.

**Абзац** – это произвольная последовательность символов, которая всегда начинается с новой строки и заканчивается нажатием клавиши Enter. Абзац может быть длиной в одну или несколько строк. Особый способ оформления абзацев – это представление их в виде списков маркированных или нумерованных.

Текстовая информация может быть представлена в виде текстового сообщения (набор символов, не всегда имеющий смысловую нагрузку) или в виде текста с присущими ему признаками текста. Признаки текста:

- Определяется тема текста.
- Выделяется главная мысль.
- Можно составить заголовок к тексту.
- Выделяются части в тексте.
- Можно установить связь между частями текста.

Для большей наглядности текстовую информацию можно представить в виде списка, таблицы, схемы или рисунка (передать смысл текста).

Для компьютерной обработки текстовой информации используются программы-приложения ОС: **текстовые редакторы**. Они позволяют выполнить следующие действия над текстовым документом:

- создавать;
- редактировать;

- форматировать;
- сохранять;
- печатать.

Текстовые редакторы имеют разную мощность. Простым, например, считается тестовый редактор Блокнот – стандартное приложение ОС Windows. Список его функций невелик. Более мощными средствами работы с текстом являются текстовые редакторы (процессоры), настольные издательские системы. Для оформления Web-страниц и сайтов используются специальные редакторы – Web-редакторы. <sup>[2]</sup> Текстовые редакторы есть и в составе систем программирования. Файлы, созданные в текстовых редакторах, имеют определенные расширения.

Рассмотрим некоторые приемы работы с текстом с помощью текстового процессора MS Word.

Текстовый документ можно создавать с помощью пустого шаблона «Новый документ» или с помощью особых шаблонов, которые уже задают структуру документа. <sup>[1]</sup>

Правила набора текста:

- Установить параметры страницы.
- Установить флажок «Линейка» через меню «Вид».
- Включить перед набором текста на панели инструментов кнопку «Непечатаемые символы» – ¶. Это маркеры пробела, неразрывного пробела, маркер абзаца, маркер конца строки.
- Между словами ставится один пробел.
- При достижении правого поля страницы текст автоматически переносится на следующую строку.
- Нажимать клавишу «Enter» нужно только в конце абзаца.
- Полезно периодически сохранять документ.
- Пробел ставится после, а не до:
  - ✓ знаков препинания;
  - ✓ закрывающейся скобки;
  - ✓ закрывающейся кавычки.
- Пробел ставится до, а не после:
  - ✓ открывающейся скобки;
  - ✓ открывающейся кавычки.

- Пробел ставится до и после тире. Перед тире ставится неразрывный (нерастяжимый) пробел: «Ctrl» + «Shift» + «пробел». Такой пробел не позволяет располагать слова на разных строчках и сохраняется фиксированным по ширине.
- Пробел не ставится:
  - ✓ после открывающейся скобки;
  - ✓ после открывающейся кавычки;
  - ✓ до и после дефиса.
- Неразрывный пробел ставится при наборе:
  - ✓ инициалов и фамилии;
  - ✓ адресов;
  - ✓ дат;
  - ✓ суммы.

Между датой и фамилией, например, ставится обычный пробел.

- В конце строки с подписью к документу ставится принудительный конец строки: «Shift» + «Enter».
- Точка в заголовке компьютерного документа не ставится. За исключением изданий для начинающих читать детей. В заголовке из двух самостоятельных предложений между ними ставят точку, а в конце опускают. Если такой заголовок не уместится в одну строку, его желательно разбить на строки так, чтобы точка попадала внутрь строки, тогда заголовок будет более четко восприниматься как единый. Остальные знаки препинания сохраняются.
- Заголовок пишется прописными буквами. Переносы слов в заголовке не разрешаются. Односложные заголовки оформляются с использованием разряженного интервала между символами. Заголовок всегда располагается на одной странице с тем текстом, к которому он относится.
- Текст от заголовка пишется через два-три интервала.
- Подзаголовки пишутся строчными буквами. Подчеркивание не используется.
- Расстояние между заголовком и подзаголовком составляет обычно два интервала.
- Заголовки и подзаголовки либо центрируются, либо выравниваются по левому краю.
- Справочная информация (Автор, название статьи, номер темы, название темы) в документе располагается в колонтитуле. Колонтитул – это часть

страницы. Он может быть только верхним, только нижним или верхним и нижним. Колонтитулы могут быть постоянными или изменяющимися по разделам документа. Если будет применяться односторонняя печать документа, то колонтитулы располагаются в одном и том же месте. Если будет применяться двусторонняя печать, то на развороте колонтитулы размещаются симметрично относительно линии соединения. В колонтитул выносятся и номер страницы.

- Колонтитулы не ставятся на титульных листах, на листах только с иллюстрациями (без текста), на листах с выходными данными издания. Верхний колонтитул не ставится на первой странице крупного раздела, а нижний – на последней.

- Используются несколько видов тире: короткое и длинное. Их можно вставить через меню «Вставка – Символ – Другие символы», вкладку «Специальные знаки». Также можно использовать «горячие клавиши»: «Ctrl» + «знак минус на цифровой клавиатуре» – это для вставки короткого тире, «Ctrl» + «Alt» + «знак минус на цифровой клавиатуре» – для вставки длинного тире. Короткое тире обычно делается шириной с букву N, отсюда и название en dash. Короткое тире часто используется при обозначении расстояний или диапазона значений. Например, « *Люди в возрасте 55–80 более подвержены стрессам*». Ширина длинного тире обычно совпадает с шириной буквы M, отсюда происходит название em dash. Длинное тире используется в роли диапазона. Например, «*Чай — с кардамоном и специями — был душист и превосходен*».<sup>[8]</sup>

- Нельзя путать символы «дефис» и «тире». Дефис используется в сложно-сочиненных словах и при переносе. Редактор автоматически переносит слова на другую строку при соответствующей настройке переносов (автоперенос).

Для препятствия нежелательного переноса слов, содержащих символ «дефис», в месте его расположения используется «неразрывный дефис»: «Ctrl» + «Shift» + «дефис».

- Для ввода некоторых символов используются таблицы символов. Нужно воспользоваться меню «Вставка» – «Символы».

- Для ввода римских цифр используются прописные латинские буквы I, V, X, L, C, D, M.<sup>[3]</sup>

**Редактирование** – это процесс внесения изменений в набранный текст, его проверка и исправление. Редактирование – это, например, выполнение следующих действий: удаление ошибочного символа, слова, строки, замена одного символа на другой, вставка пропущенных символов, слов или строк. Редактирование можно осуществлять одновременно с вводом текста. Полезно

использовать средства проверки правописания: меню «Рецензирование» – «Правописание».

**Фрагмент** – это выделенный непрерывный кусок текста, над которым можно произвести следующие операции: удаление, перемещение, копирование, форматирование. Фрагмент помещается в буфер обмена – специальную область оперативной памяти компьютера и хранится там до тех пор, пока на его место не будет помещен новый фрагмент. Фрагмент из буфера обмена можно много раз вставлять в любое место документа.

**Форматирование** текста – это изменение внешнего вида текста и его частей (страницы, абзаца, строки, символа). Форматирование применяется к выделенному фрагменту и производится с помощью панели форматирования текстового редактора. Возможно использование стилевого форматирования нескольких абзацев или символов (два вида стилей) без их предварительного выделения. **Стиль** – это определенный вид, разновидность чего-либо, выраженная в особых признаках, свойствах художественного оформления. Стиль можно создать или изменить.

**Форматирование шрифта** – это изменение параметров введенных символов (вида шрифта – гарнитуры, размера – кегля, цвета). Такое форматирование осуществляется с помощью кнопок группы команд «Шрифт» на панели форматирования редактора (диалоговое окно «Шрифт»).

**Форматирование абзаца** – задание или изменение его параметров (ширины, способа выравнивания, положения на странице, отступа в первой строке, межстрочного расстояния, интервала между абзацами). Форматирование осуществляется с помощью кнопок группы команд «Абзац» на панели форматирования редактора (диалоговое окно «Абзац»).

Параметры форматирования могут задаваться заранее на уроке – учителем, при публикации текста, например, в журнале или сборнике – издательством.

Для быстрой работы с документом в MS Word используются «горячие» клавиши (сочетание служебных клавиш и клавиш алфавитно-цифровой части клавиатуры).

Хорошим помощником является раздел «Справка» на панели основного меню текстового процессора MS Word. Справкой можно воспользоваться, используя функциональную клавишу F1.

### **Разбор решения задач**

*Задача 1.* С помощью кодировочной таблицы КОИ-8 закодируйте текстовое сообщение из трех символов «МИР». <sup>[6]</sup>

Обращаем внимание на то, что кодировочная таблица КОИ-8 не содержит строчные буквы. Ставим в соответствие каждому символу его двоичный код в таблице КОИ-8:

М – 11101101

И – 11101001

Р – 11110010

Оформляем кодовую запись сообщения, записываем коды символов друг за другом без промежутков. Получаем: 111011011110100111110010.

*Задача 2.* Декодируйте кодовую запись 041С04180440 с помощью кодировочной таблицы Unicode.<sup>[7]</sup>

Известно, что таблица Unicode двухбайтовая. Кодовая запись дана в 16-ричной системе счисления. Разделяем кодовую запись справа налево на части из 4 символов. Ставим в соответствие каждой части кодовой записи символ, используя кодовую таблицу.

041С – М

0418 – и

0440 – р

Оформляем текстовое сообщение: Мир. Последний символ «точка» – это признак конца предложения в русском языке и к декодированному сообщению он отношение не имеет.

*Задача 3.* Познакомьтесь с примером оформления маркерowanego списка в учебнике «Информатика и ИКТ 8» автора Н. Д. Угриновича, §1.1.3 «Человек и информация», в разделе «Способы восприятия информации».<sup>[2]</sup> Оформите в тетради сведения о свойствах информации, используя материал указанного параграфа, в виде маркированного списка. Свойства запишите в виде существительных. Используйте, пожалуйста, знания русского языка и литературы.

Выполненное задание может выглядеть так, как представлено далее.

Информация обладает следующими свойствами:

- понятность;
- полнота;
- точность;
- достоверность;
- актуальность;

- полезность.

Подобное задание можно выполнять и с использованием компьютера.

*Задача 4.* Наберите текст справки, используя правила набора текста. Заголовок выровнен по центру. Справка выдана Петровой Марии в том, что она учится в 8 «А» классе средней школы № 38 г. Гатчина Ленинградской области. Справка подписана директором школы И. И. Ивановым.

Задачу решаем следующим образом:

- выбираем технологию, например текстовый процессор MS Word;
- Заголовок набираем прописными буквами. Выравниваем по центру. Настройки абзаца: отступ после – 10 пт. Используем клавишу Enter – производим смену абзаца.
- Устанавливаем выравнивание текста по ширине.
- Выбираем настройки 2-го абзаца: красную (первую строку) строку с отступом на 1,25 см, междустрочный интервал – одинарный.
- Набираем текст, используя нерастяжимые пробелы. Обычный пробел ставим только после номера школы.
- Текст подписи выравниваем по правому краю.
- Сохраняем документ.

Документ будет иметь вид, представленный ниже.

#### СПРАВКА

Дана Петровой Марии в том, что она является ученицей 8 «А» класса ГБОУ школа № 38 г. Гатчина Ленинградской области.

Директор ГБОУ школа № 38 \_\_\_\_\_ И. И. Иванов

#### **Задачи для самостоятельной подготовки**

*Задача 1.* Закодировать короткое по записи существительное (на выбор) с помощью кодировочной таблицы Unicode. Подсчитать длину кода. Сравнить с длиной кода этого объекта в кодовой таблице КОИ-8 и записать обоснованный вывод.

*Задача 2.* Декодируйте следующее информационное сообщение 11101011111011111100100 с помощью кодовой таблицы КОИ-8.

*Задача 3.* Наберите текст. Все настройки стандартные. Добейтесь нужного выравнивания. Используйте служебную клавишу клавиатуры Tab. Куплет набирайте, используя сочетание клавиш {Shift+Enter}. После первого куплета вставьте пустой абзац. Наберите припев. Первая строка – висячая, с выравниванием по левому краю, с выступом. Чтобы ее установить перетащите



нижний треугольник на горизонтальной линейке вправо на нужное расстояние.  
Используйте сочетание клавиш {Shift+Enter}.

Текст расположен ниже.

## УЛЫБКА

*Музыка В. Шаинского*

*Слова М. Пляцковского*

От улыбки хмурый день светлей,  
От улыбки в небе радуга проснется...  
Поделись улыбкою своей,  
И она к тебе не раз еще вернется.

*Прпев:* И тогда наверняка, вдруг запляшут облака,  
И кузнечик запиликает на скрипке...  
С голубого ручейка начинается река,  
Ну, а дружба начинается с улыбки.  
С голубого ручейка начинается река,  
Ну, а дружба начинается с улыбки. <sup>[5]</sup>

### Литература и интернет-ресурсы

1. Информатика и ИКТ. Базовый уровень: учебник для 10 класса / Угринович Н.Д. – 6-е изд. – М. : БИНОМ. Лаборатория знаний, 2010.
2. Информатика и ИКТ. Базовый уровень: учебник для 8 класса / Угринович Н.Д. – 4-е изд. – М. : БИНОМ. Лаборатория знаний, 2011.
3. Информатика и ИКТ. Учебник для 8-9 класса. / Под ред. проф. Н.В. Макаровой. – СПб. : Питер, 2010.
4. Информатика. 6 класс: Практикум. Иванова И.А. Саратов : Лицей, 2004
5. <http://teksty-pesenok.ru/rus-v-shainskij-m-plyackovskij/tekst-pesni-ulybka/1758430/>
6. <https://ru.wikipedia.org/wiki/Windows-1251#/media/File:Ascii2.gif>
7. [https://ru.wikipedia.org/wiki/%D0%9A%D0%B8%D1%80%D0%B8%D0%B%D0%BB%D0%B8%D1%86%D0%B0\\_%D0%B2\\_%D0%AE%D0%BD%D0%B8%D0%BA%D0%BE%D0%B4%D0%B5](https://ru.wikipedia.org/wiki/%D0%9A%D0%B8%D1%80%D0%B8%D0%B%D0%BB%D0%B8%D1%86%D0%B0_%D0%B2_%D0%AE%D0%BD%D0%B8%D0%BA%D0%BE%D0%B4%D0%B5)
8. <http://inforedesign.com/notes/defis-korotkoe-tire-dlinnoe-tire/>

