

Формы представления цифровой информации Forms for presenting digital information

Аннотация

В статье рассматриваются способы машинного представления цифровой информации, а также вопросы ее реализации в памяти электронного устройства. Устанавливается взаимосвязь между различными системами счисления и рассматриваются способы восприятия информации. Сегодня цифровая информация обращается во всех сферах человеческой деятельности, без нее невозможно функционирование и развитие таких важнейших сфер деятельности человека, как строительство и энергетика, медицина и образование, искусство и культура, безопасность и охрана, политика и управление и т.д.

Ключевые слова: цифровая информация, стандартизация, множество символов, таблица кодировок, операционная система, восприятие информации, байт, десятичный код.

Annotation

The paper considers ways of machine representation of digital information, as well as issues of its implementation in the memory of an electronic device. The relationship between different number systems is established and the ways of information perception are considered. Today, digital information is used in all spheres of human activity. Without it, the functioning and development of such important spheres of human activity as construction and energy, medicine and education, art and culture, security and safety, politics and management, etc. is unthinkable.

Keywords: digital information, standardization, multiple characters, encoding table, operating system, information perception, byte, decimal code.

Воспринимать информацию могут не только люди, но и животные, и растения. Однако в отличие от людей, восприятие информации животными и растениями имеет свои особенности. Например, слоны способны воспринимать звуки, которые не слышит человек, у собак лучше всего развито обоняние, у летучих мышей – слух, а растения могут получать информацию с помощью корней и листьев. Несмотря на эти особенности, в живой природе, так же как и в мире людей, информация играет важную роль в обеспечении жизненных процессов. Воспринимаемую с помощью органов чувств информацию человек стремится выразить так, чтобы она была понятна другим. Одну и ту же информацию, в зависимости от цели деятельности, можно выразить разными способами и представить в разной форме.

По способу восприятия информации человеком можно выделить визуальную (зрительную), аудиальную (звуковую), обонятельную (запахи) вкусовую, тактильную (осязательную), вестибулярную и мышечную информацию (рис. 5).

Визуальную информацию люди воспринимают с помощью глаз. Человек может увидеть объект или явление, букву или цифру, картину или фильм, схему или карту, жест или танец.

Аудиальную информацию люди воспринимают с помощью ушей. Человек может слышать произвольные звуки, шум, музыку, пение и речь.

¹ Доцент кафедры налогового администрирования и правового регулирования Налогового института РосНОУ, кандидат экономических наук

Associate professor of the department of tax administration and legal regulation of Tax Institute RosNOU, candidate of economic sciences

Обонятельную информацию, или запахи, человек воспринимает с помощью носа. Запах можно охарактеризовать как терпкий или пряный, приятный или неприятный, тяжелый или легкий.



Рис. 5. Виды информации по способу восприятия

Вкусовую информацию человек воспринимает с помощью языка. Вкус может быть горький или сладкий, кислый или соленый.

Тактильную информацию человек воспринимает кожей. Прикасаясь к предмету, можно определить его температуру (холодный или горячий) и вид поверхности (гладкая или шероховатая, мокрая или сухая).

Вестибулярную информацию человек воспринимает с помощью вестибулярного аппарата, который отслеживает положение тела человека в трехмерном пространстве. Летя в самолете и не видя горизонта, человек может определить, куда и как он перемещается: вверх или вниз, вправо или влево, ускоренно или замедленно.

Мышечную информацию люди воспринимают с помощью мышц. Закрыв глаза, человек не пронесет ложку с супом мимо своего рта, может дотронуться указательным пальцем до своего носа, сравнить массу гири, одинаковых на ощупь.

По форме представления принято выделять числовую, текстовую, графическую, звуковую и комбинированную информацию (рис. 6).

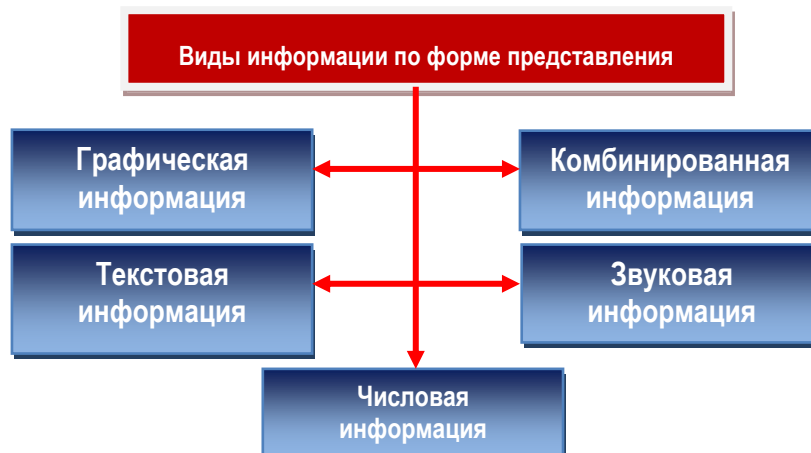


Рис. 6. Виды информации по форме представления

Например, если человек хочет выучить слова песни наизусть, то, скорее всего, он запишет стихи с помощью букв. В этом случае информация будет представлена в **текстовой** форме. Запомнить мелодию песни позволит прослушивание этой песни в исполнении певца или музыканта. В этом случае информация будет представлена в **звуковой** форме. Образ, навеянный стихами или мелодией, можно изобразить в **графической** форме с помощью рисунка.

Для того чтобы выяснить количество поклонников исполнителя песни, необходимо их подсчитать и результат представить в **числовой** форме. Каждая из этих форм представления информации имеет свои особенности. *Графическая информация* наиболее доступна, так как сразу передает визуальный образ.

С помощью *текстовой* и *звуковой информации* можно представить исчерпывающие разъяснения. *Числовая информация* дает возможность проводить различные сравнения и вычисления. Поэтому чаще всего информацию представляют в **комбинированной** форме.

Частным случаем комбинированной информации является *мультимедийная информация*, когда текстовая и числовая информация сочетается со звуковой и графической информацией, с *видеоизображением*.

Для представления информации человек использует различные *знаки*². Один и тот же знак может иметь разный смысл. Если человек наделил знак смыслом, то этот знак называют *символом*³

Например, нарисованный овал может означать или букву «О», или цифру ноль, или химический элемент кислород, или геометрическую фигуру. В нашем примере нарисованный овал – это знак. Буква, цифра и обозначение химического элемента являются символами.

Для того чтобы понимать смысл информации, представленной с помощью символов, человеку необходимо знать не только символы, но и правила составления сообщений из этих символов. Говоря другими словами, человеку необходимо знать *язык*⁴. Язык может быть *разговорным, языком рисунков, мимики и жестов, языком науки и искусства*.

Выделяют естественные (разговорные) и искусственные языки (рис. 7).

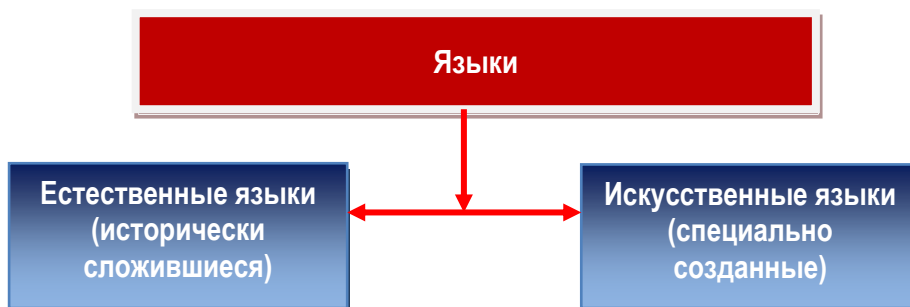


Рис. 7. Виды языков

Естественные языки исторически сложились в процессе развития человеческой цивилизации. К естественным языкам относятся русский, английский, китайский и многие другие языки. В мире насчитывается более 10 тыс. разных языков, диалектов и наречий.

² Знак – элемент конечного множества, который отличается от других элементов этого множества.

³ Символ – знак, который обладает смыслом.

⁴ Совокупность символов и правил, с помощью которых создаются осмысленные сообщения.

Искусственные языки специально созданы для профессионального применения в какой-либо области человеческой деятельности. Некоторые искусственные языки складывались в течение длительного исторического периода, например язык математических обозначений. С этой точки зрения они мало отличаются от естественных языков. Примерами искусственных языков являются эсперанто, языки программирования, язык математики, язык химии, язык логики, язык флажков на флоте, язык дорожных знаков.

Некоторые естественные языки имеют искусственно созданные алфавиты. Так, например, авторами русского языка являются Кирилл и Мефодий.

Представление информации с помощью определенного языка всегда связано с алфавитом. Алфавит⁵ содержит конечный набор символов, из которых можно составить как угодно много слов. Все символы в алфавите упорядочены. Количество символов в алфавите называют мощностью алфавита⁶.

Например, текст может быть представлен с помощью букв русского или английского алфавита, а число – с помощью алфавита десятичных цифр. В каждом из этих алфавитов буквы и цифры расположены в определенном порядке. *Мощность русского алфавита составляет 33 буквы, мощность английского алфавита – 26 букв, а мощность десятичного алфавита – 10 цифр.*

Представленную информацию можно преобразовать из одной последовательности знаков в другую, не задумываясь о смысле сообщения. Такой процесс преобразования сообщения называется *кодированием*. Обратный процесс кодированию является процесс *декодирования*. Для того чтобы выполнить кодирование или декодирование, необходимо знать правила перевода одних знаков в другие знаки. Говоря другими словами, надо знать *код*⁷ или *шифр*.

По мере развития средств появились различные способы кодирования информации.⁸ Например, кодирование с помощью азбуки (кода) Морзе (длительный сигнал – тире, короткий сигнал – точка, нет сигнала – пауза), с помощью двоичного кода (нет сигнала – 0, есть сигнал – 1). Кодирование используется для представления информации в такой форме, которая будет наиболее удобна для работы человека или технического устройства.

Например, человеку удобно и привычно работать с десятичными числами, а компьютер настроен на работу с двоичными числами. Поэтому десятичное число, введенное с помощью клавиатуры компьютера, кодируется в двоичное число. При выводе числа на экран монитора происходит декодирование из двоичного числа в десятичное число.

С точки зрения ЭВМ текст состоит из отдельных символов. К числу символов принадлежат не только буквы (заглавные или строчные, латинские или русские), но и цифры, знаки препинания, спецсимволы типа «=», «(», «&» и т.п. и даже пробелы между словами.

Обозначим основные положения кодирования информации:

- множество символов, с помощью которых записывается текст, называется *алфавитом*;

- число символов в алфавите – это его *мощность*;

⁵ Алфавит – конечный и упорядоченный набор символов, используемых для представления информации в помощью определенного языка.

⁶ Мощность алфавита – количество в нем символов.

⁷ Код – совокупность знаков и правил, предназначенных для преобразования одних знаков в другие.

⁸ Кодирование – процесс преобразования одной последовательности знаков в другую последовательность знаков.

- формула определения количества информации: $N = 2^b$, где N – мощность алфавита (количество символов); b – количество бит (информационный вес символа);
- в алфавит мощностью 256 символов можно поместить практически все необходимые символы. Такой алфавит называется *достаточным*;
- т.к. $256 = 2^8$, то вес 1 символа – 8 бит;
- единице измерения 8 бит присвоили название *1 байт*: 1 байт = 8 бит; т.е. двоичный код каждого символа в компьютерном тексте занимает 1 байт памяти;

Каким же образом текстовая информация представлена в памяти компьютера? Тексты вводятся в память компьютера с помощью клавиатуры. На клавишах написаны привычные нам буквы, цифры, знаки препинания и другие символы. В оперативную память они попадают в двоичном коде. Это значит, что каждый символ представляется 8-разрядным двоичным кодом. Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111.



Рис. 8. Перенос сигнала с клавиатуры в оперативную память компьютера

Таким образом, человек различает символы по их начертанию, а компьютер – по их коду.

Удобство побайтового кодирования символов очевидно, поскольку байт – наименьшая адресуемая часть памяти и, следовательно, процессор может обратиться к каждому символу отдельно, выполняя обработку текста. С другой стороны, 256 символов – это вполне достаточное количество для представления самой разнообразной символьной информации.

Теперь возникает вопрос, какой именно восьмиразрядный двоичный код поставить в соответствие каждому символу.

Понятно, что это дело условное, можно придумать множество способов кодировки.

Все символы компьютерного алфавита пронумерованы от 0 до 255. Каждому номеру соответствует восьмиразрядный двоичный код от 00000000 до 11111111. Этот код просто порядковый номер символа в двоичной системе счисления.

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера, называется **таблицей кодировки**. Для разных типов ЭВМ используются различные таблицы кодировки.

Международным стандартом для ПК стала таблица **ASCII** (Американский стандартный код для информационного обмена). Таблица кодов ASCII делится на две части.

Необходимо обратить внимание на то, что в таблице кодировки буквы (прописные и строчные) располагаются в алфавитном порядке, а цифры упорядочены по возрастанию значений. Такое соблюдение лексикографического порядка в расположении символов называется принципом последовательного кодирования алфавита.

Для букв русского алфавита также соблюдается принцип последовательного кодирования.

К сожалению, в настоящее время существуют пять различных кодировок кириллицы (КОИ8-Р, Windows. MS-DOS, Macintosh и ISO). Из-за этого часто возникают проблемы с переносом русского текста с одного компьютера на другой, из одной программной системы в другую.

Хронологически одним из первых стандартов кодирования русских букв на компьютерах был КОИ8 («Код обмена информацией, 8-битный»). Эта кодировка применялась еще в 70-ые годы на компьютерах серии ЕС ЭВМ, а с середины 80-х стала использоваться в первых русифицированных версиях операционной системы UNIX.

От начала 90-х годов, времени господства операционной системы MS DOS, остается кодировка CP866 («CP» означает «Code Page», «кодовая страница»).

Компьютеры фирмы Apple, работающие под управлением операционной системы Mac OS, используют свою собственную кодировку Mac.

Кроме того, Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием ISO 8859-5.

Наиболее распространенной в настоящее время является кодировка Microsoft Windows, обозначаемая сокращением CP1251.

С конца 90-х годов проблема стандартизации символьного кодирования решается введением нового международного стандарта, который называется **Unicode**. Это 16-разрядная кодировка, т.е. в ней на каждый символ отводится 2 байта памяти. Конечно, при этом объем занимаемой памяти увеличивается в 2 раза. Но зато такая кодовая таблица допускает включение до 65536 символов. Полная спецификация стандарта Unicode включает в себя все существующие, вымершие и искусственно созданные алфавиты мира, а также множество математических, музыкальных, химических и прочих символов.

Иногда бывает так, что текст, состоящий из букв русского алфавита, полученный с другого компьютера, невозможно прочитать - на экране монитора видна какая-то «абракадабра». Это происходит оттого, что на компьютерах применяется разная кодировка символов русского языка.

Список использованных источников

1. Давыдов С.Г., Логунова О.С. Проект «Индекс цифровой грамотности»: методические эксперименты. Журнал «Социология: методология, методы, математическое моделирование». 2015. № 41. С. 120-141.
2. Кучмаева О. В., Ростовская Т.К., Рязанцев С. В. Вызовы цифрового будущего и устойчивое развитие России. Социально-политическое положение и демографическая ситуация в 2017–2018 годах. М.: ИСПИ РАН, 2018
3. Лавров В. Цифровая грамотность. Секреты успешного поиска и обработки информации. – Е.: Издательские решения, 2018. 448 с.
4. Плаксин С.М., Абдрахманова Г.И., Ковалева Г.Г. Интернет-экономика в России: подходы к определению и оценке // Форсайт. 2017. Т. 11. № 1. С. 55–65.
5. Сулейманов М.Д., Бардыго Н.С., Цифровая грамотность [Текст]: учебник /. – М.: КРЕАТИВНАЯ ЭКОНОМИКА, 2019. – 324 с.