Муниципальное общеобразовательное бюджетное учреждение «Саракташская средняя общеобразовательная школа 1 имени 70-летия Победы в Великой Отечественной войне»

Исследовательская работа на тему:

**«Кодирование информации»**

Автор: Каравайцев Кирилл Алексеевич

Учащийся 10 «А» класса

МОБУ СОШ №1

Руководитель: Кузьмина Дарья Васильевна

Учитель физики МОБУ СОШ №1

**Содержание**

Введение.........................................................................................3

Актуальность…………………………………………………….4

Методы исследования...................................................................5

Результаты исследования.............................................................6

Заключение...................................................................................16

Список литературы......................................................................17

**Введение**

**Тема:** Кодирование и шифрование информации

**Цель:** Выяснить, какие существуют языки кодирования. Создать свой способ шифрования русского алфавита

**Задачи:**

1. Изучить понятия «кодирование информации» и «шифрование информации».
2. Разобрать, какие существуют способы кодирования информации.
3. Узнать зачем кодировать информацию.
4. Разработать свой способ шифрования русского алфавита.
5. Предоставить учащимся 5-8 классов несложные задачи на кодирование информации.

**Объект исследования:** Информационные шифры и коды.

**Гипотеза:** все, кто познакомится с методами кодирования информации, представленными в данном проекте, заинтересуются темой кодирования и шифрования информации и будут использовать их.

**Актуальность**

Всовременной жизничеловек сталкивается с различными способами кодирования информации. Чтобы научиться понимать информацию, представленную различными способами, необходимо ориентироваться в многообразии способов кодирования: знать и уметь интерпретировать их.

Тема проекта актуальна в связи с тем, что в 21 веке, веке информационных технологий, становятся популярны такие вакансии как программист, специалист технической поддержки и т.д. И, безусловно, ребятам, заинтересованным в этой области, необходимо изучать подобные темы.

**Методы исследования**

1. Изучение литературы и других источников информации
2. **Наблюдение**
3. **Анализ и синтез**

**Результаты исследования**

**Методы кодирования информации**

**Начнем исследование данной темы с определения “Кодирование”**

**Кодирование** – это перевод информации с одного языка на другой (запись в другой системе символов, в другом алфавите). При этом обычно кодированием называют перевод информации с «человеческого» языка на формальный, например, в двоичный код, а **декодированием** – обратный переход. Один символ исходного сообщения может заменяться одним символом нового кода или несколькими символами, а может быть и наоборот – несколько символов исходного сообщения заменяются одним символом в новом коде (так, китайские иероглифы обозначают целые слова и понятия).  Кодирование может быть равномерное и неравномерное; при равномерном кодировании все символы кодируются кодами равной длины; при неравномерном кодировании разные символы могут кодироваться кодами разной длины, это затрудняет однозначное декодирование или даже делает его невозможным.

**1.Кодирование текстовой информации в компьютере**

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

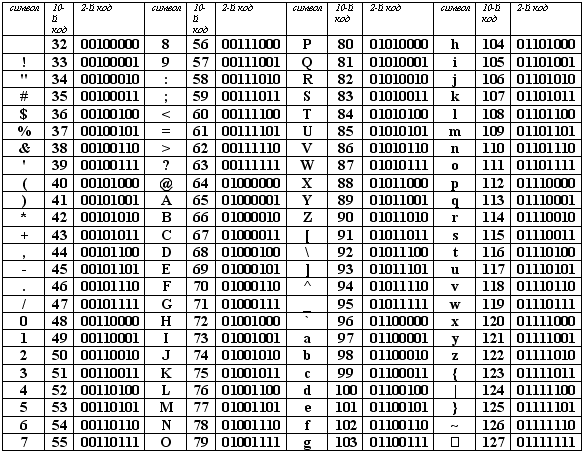
Удобство побайтового кодирования символов очевидно, поскольку байт - наименьшая адресуемая часть памяти и, следовательно, процессор может обратиться к каждому символу отдельно, выполняя обработку текста. С другой стороны, 256 символов - это вполне достаточное количество для представления самой разнообразной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

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

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.

Международным стандартом для ПК стала таблица ASCII (Американский стандартным код для информации обмена). Таблица кодов ASCII делится на две части. Международным стандартом является лишь первая половина таблицы, то есть символы с номерами от 0 (00000000), до 127 (01111111).



**2. Кодирование графической информации**

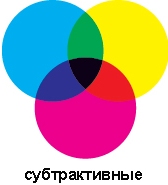
Если рассмотреть с помощью увеличительного стекла чёрно-белое графическое изображение, напечатанное в газете или книге, то можно увидеть, что оно состоит из мельчайших точек, образующих характерный узор, называемый растром. Поскольку линейные координаты и индивидуальные свойства каждой точки (яркость) можно выразить с помощью целых чисел, то можно сказать, что растровое кодирование позволяет использовать двоичный код для представления графических данных. Общепринятым на сегодняшний день считается представление чёрно-белых иллюстраций в виде комбинации точек с 256 градациями серого цвета, и, таким образом, для кодирования яркости любой точки обычно достаточно восьмиразрядного двоичного числа.

Для кодирования цветных графических изображений применяется принцип декомпозиции произвольного цвета на основные составляющие. В качестве таких составляющих используют три основные цвета: красный (Red), (Green) и синий (Blue). На практике считается, что любой цвет, видимый человеческим глазом, можно получить механического смешения этих трёх основных цветов. Такая система кодирования получила названия RGB по первым буквам основных цветов.

Такая цветовая модель называется аддитивной.

Режим представления цветной графики с использованием 24 двоичных разрядов называется полноцветным (True Color).

Каждому из основных цветов можно поставить в соответствие дополнительный цвет, т.е. цвет, дополняющий основной цвет до белого. Нетрудно заметить, что для любого из основных цветов дополнительным будет цвет, образованный суммой пары остальных основных цветов. Соответственно дополнительными цветами являются: голубой (Cyan), пурпурный (Magenta) и жёлтый (Yellow). Принцип декомпозиции произвольного цвета на составляющие компоненты можно применять не только для основных цветов, но и для дополнительных, т.е. любой цвет можно представить в виде суммы голубой, пурпурной и жёлтой составляющей.

Эта цветовая модель используется, когда имеем дело с отраженным цветом. На печатной странице цвета ведут себя иначе. Страница не излучает свет, а отражает его. При воспроизведении цветных изображений на печати мы имеем дело не со светом, а с пигментами (красками, тонером, воском), которые одни цвета поглощают, а другие отражают.

Такой метод кодирования цвета принят в полиграфии, но в полиграфии используется ещё и четвёртая краска – чёрная (Black). Поэтому данная система кодирования обозначается четырьмя буквами CMYK (чёрный цвет обозначается буквой К, потому, что буква В уже занята синим цветом), и для представления цветной графики в этой системе надо иметь 32 двоичных разряда. Такой режим также называется полноцветным.

Если уменьшить количество двоичных разрядов, используемых для кодирования цвета каждой точки, то можно сократить объём данных, но при этом диапазон кодируемых цветов заметно сокращается. Кодирование цветной графики 16-разрядными двоичными числами называется режимом High Color.

При кодировании информации о цвете с помощью восьми бит данных можно передать только 256 оттенков. Такой метод кодирования цвета называется индексным.

**3.Кодирование звуковой информации**

Приёмы и методы работы со звуковой информацией пришли в вычислительную технику наиболее поздно. К тому же, в отличие от числовых, текстовых и графических данных, у звукозаписей не было столь же длительной и проверенной истории кодирования. В итоге методы кодирования звуковой информации двоичным кодом далеки от стандартизации. Множество отдельных компаний разработали свои корпоративные стандарты, но среди них можно выделить два основных направления.

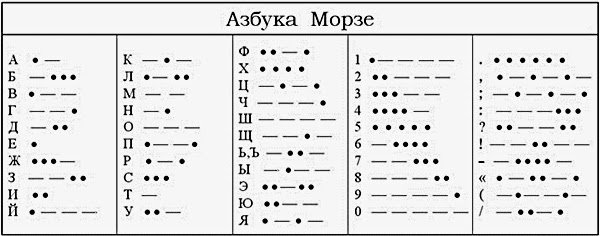
1. Метод FM (Frequency Modulation) основан та том, что теоретически любой сложный звук можно разложить на последовательность простейших гармонических сигналов разных частот, каждый из которых представляет собой правильную синусоиду, а, следовательно, может быть описан числовыми параметрами, т.е. кодом. В природе звуковые сигналы имеют непрерывный спектр, т.е. являются аналоговыми. Их разложение в гармонические ряды и представление в виде дискретных цифровых сигналов выполняют специальный устройства – аналогово-цифровые преобразователи (АЦП). Обратное преобразование для воспроизведения звука, закодированного числовым кодом, выполняют цифро-аналоговые преобразователи (ЦАП). При таких преобразованиях неизбежны потери информации, связанные с методом кодирования, поэтому качество звукозаписи обычно получается не вполне удовлетворительным и соответствует качеству звучания простейших электромузыкальных инструментов с окрасом характерным для электронной музыки. В то же время данный метод копирования обеспечивает весьма компактный код, поэтому он нашёл применение ещё в те годы, когда ресурсы средств вычислительной техники были явно недостаточны.
2. Метод таблично волнового (Wave-Table) синтеза лучше соответствует современному уровню развития техники. В заранее подготовленных таблицах хранятся образцы звуков для множества различных музыкальных инструментах. В технике такие образцы называют сэмплами. Числовые коды выражают тип инструмента, номер его модели, высоту тона, продолжительность и интенсивность звука, динамику его изменения, некоторые параметры среды, в которой происходит звучание, а также прочие параметры, характеризующие особенности звучания. Поскольку в качестве образцов исполняются реальные звуки, то его качество получается очень высоким и приближается к качеству звучания реальных музыкальных инструментов.

**Коды и шифры**

**1.Азбука Морзе**

Азбука морзе является способом знакового кодирования, представление букв алфавита, цифр, знаков препинания и других символов последовательностью сигналов: длинных (тире) и коротких (точек). За единицу времени принимается длительность одной точки. Длительность тире равна трём точкам. Пауза между элементами одного знака — одна точка, между знаками в слове — 3 точки, между словами — 7 точек.

В отличие от большинства шифров, азбука Морзе используется не для затруднения чтения сообщений, а наоборот, для облегчения их передачи (с помощью телеграфа). Длинные и короткие сигналы посылаются с помощью включения и выключения электрического тока.

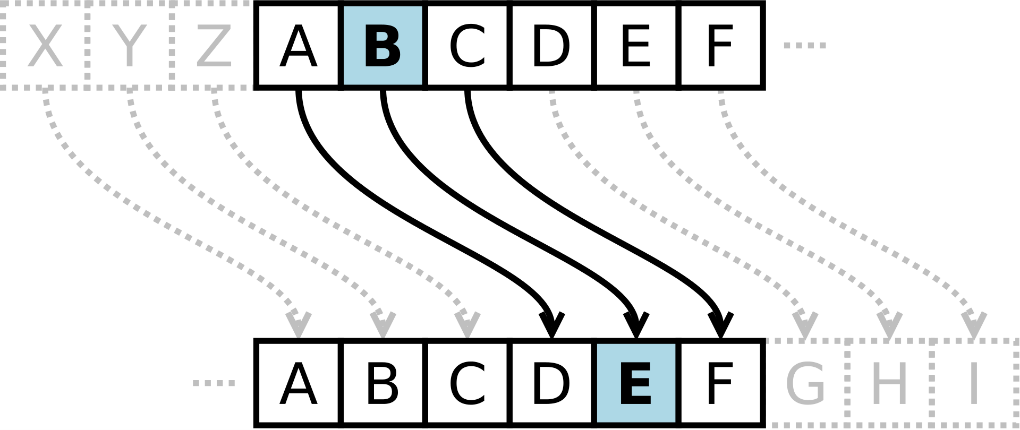


**2. Шифр Цезаря**

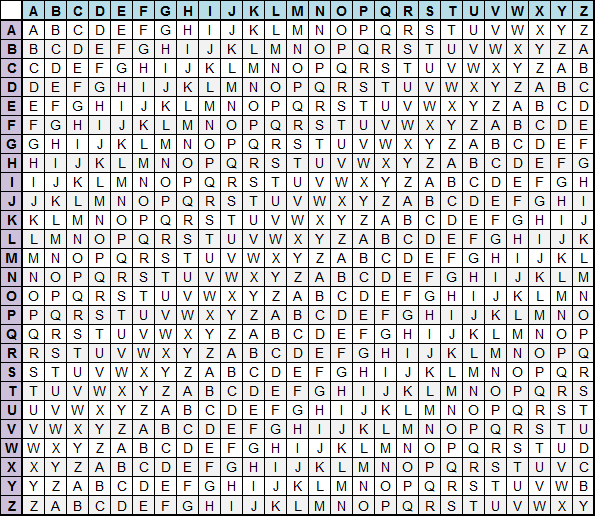
Шифр Цезаря — это вид [шифра подстановки](https://ru.wikipedia.org/wiki/%D0%A8%D0%B8%D1%84%D1%80_%D0%BF%D0%BE%D0%B4%D1%81%D1%82%D0%B0%D0%BD%D0%BE%D0%B2%D0%BA%D0%B8), в котором каждый [символ](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB) в открытом тексте заменяется символом, находящимся на некотором постоянном числе позиций левее или правее него в [алфавите](https://ru.wikipedia.org/wiki/%D0%90%D0%BB%D1%84%D0%B0%D0%B2%D0%B8%D1%82). Например, в шифре со сдвигом вправо на 3, А была бы заменена на Г, Б станет Д, и так далее.

Шифр назван в честь римского полководца Гая Юлия Цезаря, использовавшего его для секретной переписки со своими генералами.

Шаг шифрования, выполняемый шифром Цезаря, часто включается как часть более сложных схем, таких как шифр Виженера, и всё ещё имеет современное приложение в системе [ROT13](https://ru.wikipedia.org/wiki/ROT13). Как и все моноалфавитные шифры, шифр Цезаря легко взламывается и не имеет почти никакого применения на практике.



**3.Шифр Виженера**

В шифре Цезаря каждая буква алфавита сдвигается на несколько позиций; например в шифре Цезаря при сдвиге +3, A стало бы D, B стало бы E и так далее. Шифр Виженера состоит из последовательности нескольких шифров Цезаря с различными значениями сдвига. Для зашифровывания может использоваться таблица алфавитов, называемая tabula recta или квадрат (таблица) Виженера. Применительно к латинскому алфавиту таблица Виженера составляется из строк по 26 символов, причём каждая следующая строка сдвигается на несколько позиций. Таким образом, в таблице получается 26 различных шифров Цезаря. На каждом этапе шифрования используются различные алфавиты, выбираемые в зависимости от символа ключевого слова.

Очень долгое время шифр Виженера считался невзламываемым. Чтобы его расшифровать, для начала угадывают длину кодового слова и применяют частотный анализ к каждой n-ной букве послания, где n — предполагаемая длина кодового слова. Если длина была угадана верно, то и сам шифр вскроется с большей или меньшей долей вероятности. Если предполагаемая длина не дает верных результатов, то пробуют другую длину кодового слова, и так далее до конца.

**4. Шифр Тритемиуса**

Шифр Тритемиуса — система шифрования, разработанная [Иоганном Тритемием](https://ru.wikipedia.org/wiki/%D0%A2%D1%80%D0%B8%D1%82%D0%B5%D0%BC%D0%B8%D0%B9,_%D0%98%D0%BE%D0%B3%D0%B0%D0%BD%D0%BD). Представляет собой усовершенствованный [шифр Цезаря](https://ru.wikipedia.org/wiki/%D0%A8%D0%B8%D1%84%D1%80_%D0%A6%D0%B5%D0%B7%D0%B0%D1%80%D1%8F), то есть [шифр подстановки](https://ru.wikipedia.org/wiki/%D0%A8%D0%B8%D1%84%D1%80_%D0%BF%D0%BE%D0%B4%D1%81%D1%82%D0%B0%D0%BD%D0%BE%D0%B2%D0%BA%D0%B8). По алгоритму шифрования, каждый [символ](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB) сообщения смещается на символ, отстающий от данного на некоторый шаг. Здесь шаг смещения делается переменным, то есть зависящим от каких-либо дополнительных факторов. Например, можно задать закон смещения в виде линейной функции (уравнения зашифрования) позиции шифруемой буквы. Сама функция должна гарантировать целочисленное значение. Прямая функция шифрования должна иметь обратную функцию шифрования, тоже целочисленную.

Формула для шифра: L = (m+k) mod N L-номер зашифрованной буквы в алфавите, m-порядковый номер буквы шифруемого текста в алфавите, k-число сдвига, N-количество букв в алфавите. Число сдвига (шаг смещения) в данном случае переменная величина, функционально зависящая от позиции символа в сообщении. Зависимость может быть разной, например, линейной или квадратичной. k = A\*p + B (линейная зависимость). A, B – ключи. p – позиция буквы в сообщении.

**5.Хэш-код MD5**

Данный код популярен в нынешнее время информационных технологий и кибербезопасности. Широко применяется для проверки целостности информации и хранения хешей паролей.

**MD5** является одним из алгоритмов хеширования на 128-битной основе. Под хешированием понимают преобразование входных данных по определенному алгоритму в битовую строку определенной длины. При этом полученный в ходе вычислений результат представлен в шестнадцатеричной системе исчисления. Она называется хешем, хеш-суммой или хеш-кодом.

Процесс хеширования широко применяется в программировании и веб-индустрии. В основном для создания уникальных значений в ассоциативных массивах, идентификаторов.

Область применения хеш-кодов:

* Создание электронных подписей;
* Хранение паролей в базах данных систем безопасности;
* В рамках современной криптографии для создания уникальных ключей онлайн;
* Проверка подлинности и целостности элементов файловой системы ПК.

**MD5** как стандарт хеширования был разработан в 1991 году для создания уникального хеш-кода от заданного значения с последующей проверкой его подлинности.

**Заключение**

Вся жизнь человека, так или иначе, связана с получением, накоплением и обработкой информации. Чтобы человек ни делал: читает ли он книгу, смотрит ли он телевизор, разговаривает, он постоянно и непрерывно получает и обрабатывает информацию. Для любой операции над информацией (даже такой простой, как сохранение) она должна быть как-то представлена (записана, зафиксирована). Этот процесс имеет специальное название – кодирование информации. Современному человеку невозможно обойтись без кодирования информации. Чтобы мы понимали все из того, что слышим, видим, чувствуем необходимо знать различные способы представления (кодирования) информации.

Я пришел к выводу: криптография как наука о шифровании, основанная на способах кодирования информации, востребована в наше время и будет востребована в будущем. Так как без кодирования сейчас не обходиться ни одно государство, ни один банк, ни одно предприятие.

Работая над проектом, я узнал много нового об уникальных способах шифрования. Надеюсь и ознакомившиеся с ними люди, тоже откроют для себя много интересного.

**Список литературы**

1. [https://ru.wikipedia.org/wiki/Криптография#Современная\_криптография](https://ru.wikipedia.org/wiki/Криптография%23Современная_криптография)
2. [https://xakep.ru/2016/06/30/crypto-part4/#toc01.](https://xakep.ru/2016/06/30/crypto-part4/%23toc01.)
3. <https://tproger.ru/translations/10-codes-and-ciphers/>
4. [https://ru.wikipedia.org/wiki/MD5#Алгоритм\_MD5](https://ru.wikipedia.org/wiki/MD5%23Алгоритм_MD5)
5. <https://www.internet-technologies.ru/articles/newbie/heshirovanie-i-rasshifrovka-md5-hesh-koda.html>
6. <https://ru.wikipedia.org/wiki/Шифр_Тритемиуса?veaction=edit>
7. <https://it-black.ru/populjarnye-kody-i-shifry/>