

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
**Федеральное государственное автономное образовательное учреждение**  
**высшего образования**  
**«Севастопольский государственный университет»**

**ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**  
**КОДИРОВАНИЯ, СЖАТИЯ И ШИФРОВАНИЯ ДАННЫХ**  
**(ТЕХНОЛОГИИ ОПТИМАЛЬНОГО КОДИРОВАНИЯ)**

Методические указания к выполнению лабораторной работы №1  
по дисциплине «Информационные технологии»

для студентов очной и очно-заочной форм обучения, обучающихся по  
направлению 09.03.02 Информационные системы и технологии

**Севастополь**  
**2015**

**Исследование информационных технологий кодирования, сжатия и шифрования данных (технологии оптимального кодирования).**

Методические указания к выполнению лабораторной работы №1 по дисциплине «Информационные технологии»/ Сост. И.В. Кудрявченко. – Севастополь: Изд-во СевГУ, 2015. – 26с.

Целью методических указаний является обеспечение наглядного изучения методов и алгоритмов оптимального кодирования сообщений, применяемых в информационных технологиях передачи и хранения данных, а так же приобретение студентами устойчивых практических навыков применения методов оптимального (эффективного) кодирования данных.

Методические указания предназначены для студентов направления подготовки 09.03.02 очной и очно-заочной форм обучения

Методические указания рассмотрены и утверждены на заседании кафедры информационных систем (протокол №1 от 31 августа 2015 г.)

Допущено учебно-методическим центром СевГТУ в качестве методических указаний

Рецензент: Карлусов В.Ю., к.т.н., доцент кафедры информационных систем

## СОДЕРЖАНИЕ

1 Цель работы.....	3
2 Краткие теоретические сведения.....	3
3 Порядок выполнения исследования.....	21
4 Содержание отчета о выполнении лабораторной работы.....	21
5 Контрольные вопросы.....	22
Библиографический список.....	23
Приложение А. Варианты заданий.....	24

## Лабораторная работа №1 (4ч)

«Исследование технологий оптимального кодирования»  
(на выполнение работы отводится 8 часов, из них 4 часа СРС)

### 1 ЦЕЛЬ РАБОТЫ

Изучение особенностей методов статического сжатия данных и алгоритмов простого безизбыточного и оптимального кодирования (Шеннона-Фано, Хаффмана, арифметического), а также приобретение практических навыков информационных расчетов эффективности работы кодеров/декодеров систем передачи и хранения данных.

### 2 КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

#### 2.1 Общие положения

##### 2.1.1 Информационные технологии

*Определение.* Информационная технология — это совокупность средств и методов их применения для целенаправленного изменения свойств информации, определяемого содержанием решаемой задачи или проблемы.

Каждое информационное преобразование (ИП) характеризуется содержанием, направлением и объемом. Содержание ИП определяется конкретным набором изменяемых свойств информации.

В информационных технологиях применяются следующие ИП: сбор, накопление, регистрация, передача, копирование, упорядочение, хранение, поиск, представление, выдача и защита информации. В большинстве из перечисленных ИП используются методы и алгоритмы кодирования, сжатия и шифрования данных.

##### 2.1.2 Структурная схема системы передачи данных

На рисунке 2.1 изображена обобщенная структурная схема системы передачи данных (СПД). Эта схема применяется для анализа информационных систем временной и пространственной передачи данных.

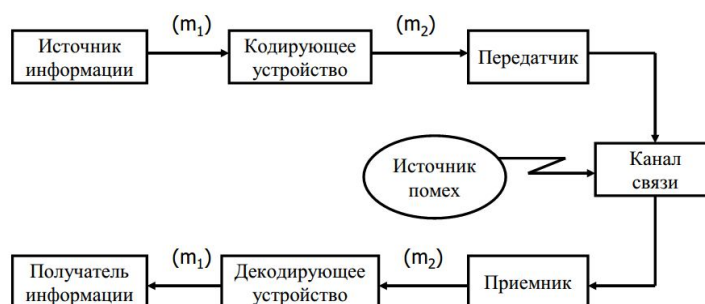


Рисунок 2.1 — Структурная схема системы передачи данных

**Источник информации (ИИ).** Сообщения, вырабатываемые ИИ, могут иметь различную физическую природу: цифровые данные, полученные от ЭВМ, речь, изображения, команды управления, результаты измерения физических величин. Если необходимо передавать информацию об аналоговых

величинах, то предварительно можно преобразовать аналоговый сигнал в цифровой, используя аналого-цифровой преобразователь (АЦП). Это позволяет рассматривать ИИ в качестве дискретного источника, т.е. полагать, что передаче подлежит одно или несколько сообщений, выбранных из некоторого множества возможных сообщений, а ИИ вырабатывает символы (сигналы) *первичного алфавита*. Мощность первичного алфавита обозначим  $m_1$ . В дальнейшем будем исходить из того, что ИИ является дискретным.

**Кодирующее устройство (КУ).** Для передачи информации по каналу связи используют сигналы с небольшим числом однозначно различимых качественных признаков  $m_2$ . Такими признаками являются: значения амплитуды, частоты, фазы сигнала или сочетания указанных параметров. Обычно число различимых качественных признаков сигналов невелико и не превышает мощности первичного алфавита

$$m_1 > m_2. \quad (2.1)$$

Если условие (2.1) выполняется, то КУ является обязательным элементом СПД. В результате кодирования на выходе КУ образуются символы (сигналы), составляющие *вторичный алфавит*. В случае применения двоичного кодирования количество качественных признаков сигнала равно двум ( $m_2=2$ ).

**Передатчик (Прд).** Прд преобразует закодированное сообщение в сигналы, которые соответствуют характеристикам применяемого КС, т.е. КУ и Прд выполняют *согласование ИИ с КС*. Например, видеоимпульсы (последовательности логических нулей и единиц) не могут непосредственно передаваться по радиоканалам или каналам оптической связи. В этом случае, в составе Прд должен находиться модулятор, который переносит спектр исходного сигнала в рабочий диапазон частот применяемого КС. Обычно в СПД применяются модуляторы на основе амплитудной, фазовой, частотной модуляции (манипуляции).

**Канал связи.** КС — среда, используемая для передачи сигнала от Прд к приемнику. Термин «канал связи» не является синонимом «линии связи».

**Приемник (Прм).** Прм выполняет преобразования, обратные тем, которые выполняет Прд.

**Декодирующее устройство (ДУ).** ДУ выполняет преобразования, обратные тем, которые выполняет КУ.

**Получатель информации (ПИ).** ПИ — человек или техническое устройство, для которого предназначено переданное сообщение.

**Источник помех (ИП).** В процессе передачи по КС сигнал может искажаться помехами (шумами). Для учета действия помех (шумов) в схему СПД добавлен ИП. Помехи (шумы) считают сосредоточенными в КС. Математическое описание КС задается в виде вероятностных характеристик сигнала на стороне Прд и Прм. Влияние ИП на передаваемые по КС сообщения проявляется в виде искажений принимаемых символов.

### Закон распределения ошибок в КС

Если ошибки в передаваемом сообщении происходят независимо друг от друга, то их распределение подчиняется биномиальному закону

$$P = \sum_{i=0}^n C_n^i p_z^i (1-p_z)^{n-i}, \quad (2.2)$$

где

$n$  — общее количество элементов в сообщении;

$i$  — кратность ошибки в последовательности;

$p_z$  — вероятность искажения одного элемента;

$1-p_z$  — вероятность правильной передачи элемента;

$C_n^i = \frac{n!}{i!(n-i)!}$  — число сочетаний из  $n$  по  $i$ .

### 2.1.3 Кодирование

*Определения*

1. Кодирование — однозначное преобразование символов первичного алфавита в символы вторичного.
2. Код представляет собой алгоритм, по которому осуществляется кодирование.

Необходимость кодирования объясняется схемой на рисунке 2.1.

Современные методы оптимального, помехоустойчивого кодирования и шифрования позволяют решать следующие задачи:

- выполнять согласование ИИ с КС и минимизировать длину передаваемых сообщений;
- обеспечивать помехоустойчивую передачу данных по КС;
- обеспечить конфиденциальность передаваемых данных.

Иными словами, с помощью кодирования осуществляется согласование параметров КС и передаваемых сообщений, что позволяет более экономно использовать полосу частот КС, а также уменьшать стоимость передачи и хранения сообщений.

### 2.1.4 Информационные параметры ИИ и КС

Основными информационными параметрами ИИ и КС являются: количество информации ( $I$ ); энтропия ( $H$ ), объем информации ( $Q$ ) пропускная способность КС ( $C$ ), скорость передачи информации ( $R$ ).

#### Количество информации.

Для *равновероятного* алфавита количеством информации, определяется в соответствии с мерой Хартли:

$$I = \log_2 N = \log_2 m^n = n \log_2 m, \quad (2.3)$$

где  $N$  - общее количество передаваемых сообщений;

$m$  - мощность первичного алфавита;

$n$  - число символов первичного алфавита в сообщении.

Если в (2.3) основание логарифма равно двум, то количество информации измеряется в *битах*. Если применяют десятичный логарифм, то единицей измерения количества информации является *дит*. Если применяют натуральный логарифм, то единицей является *нат*.

Ограничения Хартли:

- рассматриваются только дискретные сообщения;
- множество различных сообщений конечно;
- символы, составляющие сообщения, равновероятны и независимы.

2. Для *неравновероятного* алфавита ИИ количество информации определяется по формуле Шеннона

$$I = n \sum_{i=1}^m p_i \log_2 \frac{1}{p_i}, \quad (2.4)$$

где  $p_i$  — вероятность появления  $i$ -го символа в сообщении. Остальные обозначения соответствуют формуле (2.3).

### Энтропия

Энтропия — удельное количество информации (количество информации, приходящейся на одну букву сообщения). Энтропия является мерой неопределенности. Если в сообщении имеется  $n$  элементов, то

$$H = \frac{I}{n} = - \sum_{i=1}^m p_i \log_2 p_i = \sum_{i=1}^m p_i \log_2 \frac{1}{p_i}. \quad (2.5)$$

### Объем информации

Объем информации равен произведению средней длины сообщения во вторичном алфавите на число сообщений

$$Q = kn, \quad (2.6)$$

где  $k$  — средняя длина сообщения во вторичном алфавите;

$n$  — общее число сообщений.

Объем информации определяется относительно вторичного алфавита и не зависит от числа повторений.

Количество информации не м.б. больше объема той же информации в соизмеримых единицах.

Объем информации равен ее количеству, если выполняются следующие условия:

- для равновероятных алфавитов количество символов первичного алфавита является целой степенью  $m_2$ ;
- для неравновероятных алфавитов  $p_i = m_2^{-n_i}$ , где  $n_i$  — целое число.

### Скорость передачи информации

Скорость передачи информации характеризует количество информации, приходящееся в среднем на одно сообщение.

Если в секунду передается  $n$  сообщений, то скорость передачи информации будет составлять

$$R=n \cdot [H(X)-H(X/Y)] \quad (2.7)$$

где

$I(X,Y)=H(X)-H(X/Y)$  – взаимная информация, которую определяют как уменьшение энтропии системы  $X$  (ИИ) в результате получения сведений о системе  $Y$  (принятое сообщение);

$H(X)$  — априорная энтропия до наблюдения;

$H(X/Y)$  — остаточная энтропия после получения сведений;

$I(X,Y)$  — полная информация о системе  $X$ , содержащаяся в системе  $Y$ .

### Пропускная способность КС

*Определения*

1. Пропускная способность КС есть максимально достижимая для данного КС скорость передачи информации

$$C=\max R=n \cdot [H(X)-H(X/Y)] \max \quad (2.8)$$

2. Пропускная способность КС есть максимальное количество информации, которое может быть передано через КС за единицу времени

$$C=nI(X,Y) \max \quad (2.9)$$

### Теорема Шеннона 1

Пусть имеется ИИ с энтропией  $H(X)$  и канал без шума (КБШ) с пропускной способностью  $C$ .

Если  $C > H(X)$ , то всегда можно закодировать достаточно длинное сообщение таким образом, что оно будет передано без задержек.

Если  $C < H(X)$ , то передача информации без задержек невозможна.

В любом реальном КС присутствуют помехи. Однако, если уровень помех мал и вероятность искажения практически равна нулю, то полагают, что все сигналы передаются неискаженными. В этом случае среднее количество информации, переносимое одним символом, равно  $I(X,Y)=I(X,X)=H(X)$ . Максимальное значение энтропии  $H_{\max}=\log(m)$ . Следовательно, пропускная способность дискретного КБШ за единицу времени равна  $C=n \cdot \log(m)$ .

### Теорема Шеннона 2

Пусть имеется ИИ  $X$ , энтропия которого в единицу времени равна  $H(X)$ , и КС с шумами (КСШ) с пропускной способностью  $C$ .

Если  $H(X) > C$ , то при любом кодировании передача сообщений без задержек и искажений невозможна.

Если  $H(X) < C$ , то любое достаточно длинное сообщение всегда можно закодировать так, что оно будет передано без задержек и искажений с вероятностью сколь угодно близкой к единице.

### Основное условие согласования ИИ и КС

Для наиболее эффективного использования КС необходимо, чтобы скорость передачи информации была как можно ближе к пропускной способности КС.

$$C \geq R. \quad (2.10)$$

Неравенство (2.10) задает основное условие согласования ИИ и КС.

### Скорость передачи информации в непрерывном КС

Если применяется непрерывный КС, то соотношение (2.10) также выполняется, но максимальная скорость передачи определяется по формуле

$$R_{\max} = F_m \log_2 \left( 1 + \frac{P_c}{P_{\text{ш}}} \right), \quad (2.11)$$

где

$F_m$  — максимальная полоса частот КС;

$P_c$  — средняя мощность сигнала;

$P_{\text{ш}}$  — средняя мощность шума или помехи с равномерным спектром (тепловой «белый шум» — БШ).

Анализируя выражение (2.11), можно сделать следующие выводы:

1. Полоса пропускания КС должна быть тем шире, чем шире спектр передаваемого сигнала и чем меньше допустимое в КС отношение сигнал/шум.
2. Применяя достаточно сложные системы кодирования можно передавать двоичные знаки с максимальной скоростью (2.11) при сколь угодно малой частоте ошибок (частота ошибок  $K_{\text{ош}} = N_{\text{ош}}/N$ , где  $N$  — общее число сообщений,  $N_{\text{ош}}$  — количество ошибочно переданных сообщений).
3. Для достижения предельной скорости (2.11) передаваемые сигналы должны приближаться по своим статистическим свойствам к БШ.

### 2.1.5 Избыточность информации

Известно, что максимальное количество информации на символ сообщения можно получить только в случае равновероятных и независимых символов. Реальные коды обычно не удовлетворяют этому условию в полной мере. Поэтому вводят понятие информационной избыточности сообщения.

Избыточность характеризует *информационную недогруженность* символа сообщения.

Различают *естественную* и *искусственную* избыточность.

Естественная избыточность относится к первичным алфавитам (либо заложена в структуре сообщения), а искусственная — к вторичным.

Естественная избыточность бывает *семантической* и *статистической*.

Семантическая избыточность связана с тем, что мысль, высказанная в сообщении может быть сформулирована более кратко. Устраняют семантическую избыточность в первичном алфавите (аббревиатуры, условные сокращения часто повторяющихся слов).



Статистическая избыточность связана с тем, что символы ИИ имеют разную вероятность появления в сообщении. Некоторые ИИ могут иметь взаимозависимые вероятности появления символов (источники с памятью).

Сообщения на естественных языках, используемые для передачи информации, можно значительно сжать.

Для оценки сжатия сообщений (данных) применяют коэффициент сжатия

$$\mu = K_{\text{ск}} = \frac{H}{H_{\text{max}}}, \quad (2.12)$$

где

$H$ —текущая энтропия;

$H_{\text{max}} = \log_2 m$  — максимальная энтропия, определяемая при равновероятном появлении символов в сообщении.

Избыточность сообщения определяют по формуле (2.13)

$$D = 1 - \frac{H}{H_{\text{max}}} = 1 - \mu. \quad (2.13)$$

### Основная теорема кодирования для КБШ

При кодировании множества сигналов с энтропией  $H$  в алфавите, насчитывающем  $m$  символов, при условии отсутствия шумов, средняя длина кодового слова не может быть меньше, чем  $H/(\log m)$ .

Если вероятности символов не являются отрицательными степенями числа  $m$ , то точное достижение границы  $H/(\log m)$  невозможно, но при кодировании достаточно длинными блоками к этой границе можно сколь угодно приблизиться.

Обозначим количество независимых букв в блоке  $M$ , а среднюю длину кодового слова данного алфавита  $L$ . Тогда

$$\frac{H}{\log m} \leq L < \frac{H}{\log m} + \frac{1}{M}. \quad (2.14)$$

Из соотношений (2.14) следует:

- среднее число двоичных знаков на букву в точности равно энтропии источника сообщений, если кодируемый алфавит равновероятный и  $m=2^i$  ( $i = 1, 2, 3, \dots$ ) или вероятности появления символов сообщения являются целочисленными отрицательными степенями двойки ( $p_i=2^{-i}$ );
- чем длиннее первичное кодовое слово, тем точнее величина  $H/(\log m)$  характеризует среднюю длину кодового слова;
- чем больше длина блока, тем меньше разность между верхней и нижней границами, определяющими среднее число элементарных символов на букву сообщения;
- из какого бы числа букв не состоял алфавит, целесообразно кодировать сообщения не побуквенно, а поблочно;
- энтропия первичного алфавита может характеризовать возможный предел сокращения кодового слова во вторичном алфавите.

## 2.2 Оптимальное (эффективное) кодирование

### 2.2.1 Общие положения

*Определение.* Оптимальным (эффективным) кодированием называется преобразование символов первичного алфавита  $m_1$  в кодовые слова во вторичном алфавите  $m_2$ , при котором средняя длина сообщений во вторичном алфавите имеет минимально возможную для данного  $m_2$  длину.

Оптимальное кодирование предполагает, что помехи в КС отсутствуют или минимальны. Преимущество оптимальных кодов состоит в том, что они позволяют увеличивать скорость передачи сообщений и уменьшают расход памяти.

#### Свойства оптимальных кодов

1. Минимальная средняя длина кодового слова оптимального кода обеспечивается в том случае, когда избыточность каждого кодового символа сведена к минимуму (в идеальном случае — к нулю).
2. Кодовые слова оптимального кода должны строиться из равновероятных и взаимонезависимых символов.

Все оптимальные коды делятся на *равномерные* и *неравномерные*.

К равномерным кодам относят двоичные позиционные коды, любые  $m$ -значные позиционные коды и коды, построенные на их основе.

Примерами равномерных эффективных кодов являются двоичный безизбыточный код ( $m_2=2$ ) в котором разрешены все кодовые комбинации и код Грея (КГ).

#### Равномерные эффективные коды

##### Код Грея и двоичный безизбыточный код

Код Грея (КГ) — рефлексный или отраженный двоичный код, представляющий двоичную систему нумерации, в которой два соседних значения различаются только в одном двоичном разряде.

КГ был разработан для защиты электромеханических переключателей от ложного срабатывания. Впоследствии КГ стали применять для выявления и исправления ошибок в системах связи и телеуправления.

Название рефлексный (отражённый) двоичный код связано со структурой КГ: комбинации второй половины значений КГ эквивалентны комбинациям первой половины, расположенным в обратном порядке (за исключением старшего бита, который просто инвертируется).

Т.к. подавляющее большинство телемеханических объектов имеет плавные характеристики, то и для передаваемой телемеханической информации характерен плавный переход от одного значения к другому. Если при использовании КГ в принятом сообщении одновременно изменяется несколько символов, то это свидетельствует о помехах в КС или о нарушении режима работы контролируемого объекта. В частности, КГ может применяться для исправления ошибок преобразования аналоговых сигналов в цифровые.

КГ часто используются в датчиках угла поворота, а также для кодирования номера дорожки жесткого диска.

### Пример 1

Пусть  $n=3$ . Тогда двоичный безизбыточный код (ДК) имеет 8 разрешенных комбинаций, которые образуют последовательность:

000 – 001 – 010 – 011 – 100 – 101 – 110 – 111,

соответствующую последовательности десятичных чисел:

0 – 1 – 2 – 3 – 4 – 5 – 6 – 7.

Для защиты электромеханических переключателей от ложного срабатывания ДК не годится, т.к. в нем имеются соседние значения, для которых предыдущая комбинация отличается от последующей всеми разрядами (3 — 4: 011 — 100).

Для той же последовательности десятичных чисел трехразрядный код Грея образует последовательность: 000 — 001 — 011 — 010 — 110 — 111 — 101 — 100, в которой соседние числа различаются только в одном разряде.

Известны правила перевода ДК в КГ и обратно.

#### Прямой перевод (ДК → КГ)

1. Под двоичным кодом записывают то же число со сдвигом вправо на один разряд (младший разряд при этом теряется).
2. Выполняют поразрядное сложение сдвинутого и не сдвинутого чисел по модулю два.

#### Обратный перевод (КГ → ДК)

1. Цифра старшего разряда остается без изменений.
2. Каждая последующая цифра инвертируется столько раз, сколько единиц ей предшествует в КГ.

### Неравномерные эффективные коды

#### Основные принципы оптимального кодирования

1. В неравномерных кодах наиболее вероятным символам первичного алфавита присваиваются более короткие кодовые комбинации, а символам с меньшей вероятностью — более длинные. Это способствует увеличению скорости передачи информации по КС.
2. Выбор каждого кодового слова необходимо производить так, чтобы содержащееся в нем количество информации было максимальным.

Примерами неравномерных эффективных (оптимальных) кодов являются коды Шеннона-Фано, Хаффмана и арифметический код.

#### Код Шеннона-Фано

Построение КШФ для двоичных алфавитов ( $m=2$ )

1. Все символы первичного алфавита ИИ располагают в ряд сверху вниз в порядке убывания их вероятностей.
2. Упорядоченные символы делят на две группы с примерно равными суммарными вероятностями. Одной из групп присваивают код «1», второй – код «0».
3. Затем каждую из подгрупп снова делят на две группы с примерно равными вероятностями. Им также присваиваются коды «1» и «0». Такое разбиение продолжают до тех пор, пока в каждой подгруппе не останется по одному символу.

4. Далее кодируют символы. Для каждого символа записывают коды подгрупп, в которые он входит, начиная от первого до последнего деления.

### Пример 2

Требуется закодировать сообщение АВАВАСАD кодом Шеннона-Фано (ШФ) и двоичным безизбыточным кодом (ДК).

Построение кода ШФ поясняет таблица 2.1, в которой также имеется кодовая таблица ДК. Результаты кодирования сравниваются в таблице 2.2.

Таблица 2.1 – Кодовые таблицы (ДК, ШФ) и схема кодирования ШФ

Символы	Кодовые слова ДК	Частоты символов	1 деление	2 деление	3 деление	Кодовые слова КШФ
А	00	0,5	0	—		0
В	01	0,25	1	0	—	10
С	10	0,125		1	0	110
Д	11	0,125			1	111

Таблица 2.2 – Результаты кодирования

Исходное сообщение	Закодированное сообщение	Длина сообщения
АВАВАСАD	00 01 00 01 00 10 00 11 А В А В А С А D	16 бит 11/5 («0»/«1»)
	0 10 0 10 0 110 0 111 А В А В А С А D	14 бит 7/7 («0»/«1»)

Расчет средней длины в битах символа сообщения (КШФ)

$$L_{\text{cp}} = \frac{N_2}{N_1} = \frac{14}{8} = 1,75 \left[ \frac{\text{бит}}{\text{симв}} \right] \quad (2.15a)$$

$$L_{\text{cp}} = \sum_{i=1}^{m1} p(a_i) l(a_i) = 0,5 \cdot 1 + 0,25 \cdot 2 + 0,125 \cdot 3 + 0,125 \cdot 3 = 1,75 \left[ \frac{\text{бит}}{\text{симв}} \right], \quad (2.15б)$$

где  $N_1$  – количество символов в исходном сообщении;

$N_2$  – количество бит в закодированном сообщении;

$m1$  – мощность первичного алфавита.

Расчет энтропии сообщения

$$H(A) = \sum_{i=1}^{m1} p(a_i) \log \frac{1}{p(a_i)} = 0,5 \cdot \log 2 + 0,25 \cdot \log 4 + 0,125 \cdot \log 8 + 0,125 \cdot \log 8 = 1,75 \left[ \frac{\text{бит}}{\text{симв}} \right].$$

### Расчет избыточности сообщения

$$D = 1 - \frac{H}{H_{\max}} = 1 - \frac{1,75}{2} = 0,125.$$

### Выводы

1. В результате кодирования по методу ШФ получено сообщение с нулевой избыточностью, т.к. достигнута нижняя граница соотношения (2.14). Из таблицы 2.2 видно, что количества нулей и единиц в сжатом сообщении совпадают.
2. Двоичное безизбыточное кодирование обеспечивает среднюю длину символа сообщения 2 [бит/симв] и сохраняет избыточность исходного сообщения.
3. Код ШФ является неоднозначным, т.к. результат кодирования в общем случае зависит от порядка группирования символов в подгруппы с примерно равными суммарными вероятностями.

### Статический код Хаффмана

Методика построения статического кода Хаффмана (КХ) является однозначной.

1. Все символы первичного алфавита располагают в ряд сверху вниз в порядке убывания их вероятностей.
2. Два последних символа объединяются в один вспомогательный символ, вероятность которого равна сумме вероятностей объединяемых символов.
3. Строится не возрастающий ряд со вспомогательным символом.
4. Повторяются п.п. 2 и 3 данной методики до тех пор, пока не останется один вспомогательный (суммарный) символ с вероятностью 1.
5. Строится дерево Хаффмана, корнем которого является суммарный символ. Ветвление дерева начинают от суммарной вершины. Ветви с большей вероятностью присваивают код 1 (правой), а ветви с меньшей — код 0 (левой).
6. Тупиковые вершины (листья) — исходные символы.
7. Кодовые комбинации записывают от корня до тупиковой вершины.

### Пример 3

Требуется закодировать сообщение АВАВАСАД с помощью статического КХ.

Порядок построения кода поясняет таблица 2.3 и дерево на рисунке 2.2. Кодовые слова в таблице 2.3 записаны в соответствии с построенным деревом Хаффмана.

Таблица 2.3 – Кодовая таблица КХ

Символы	Частоты символов				Кодовые слова КХ
	1	2	3	4	
A	0,5	0,5	0,5	1(ABCD)	1
B	0,25	0,25	0,5(BCD)		01
C	0,125	0,25 (CD)			001
D	0,125				000

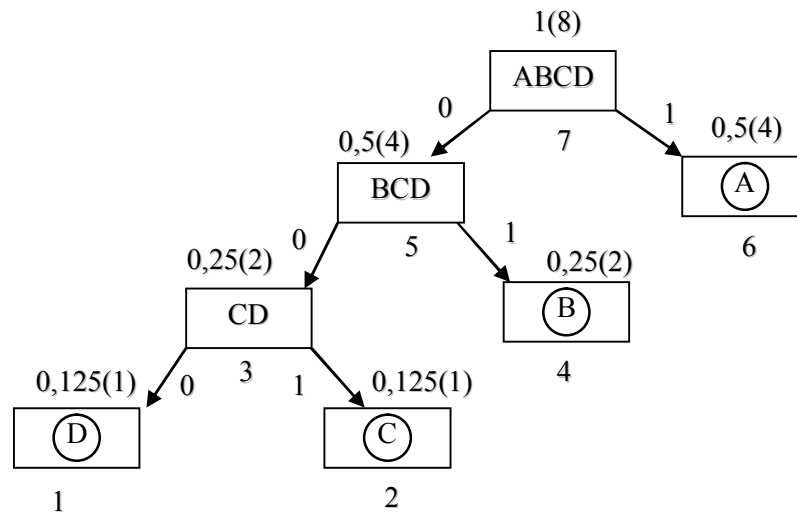


Рисунок 2.2 – Дерево Хаффмана

Статический код Хаффмана является двухпроходным: вначале анализируется текстовый файл, составляется кодовая таблица по вышеописанной методике, после чего сообщение кодируется и передается по каналу связи декодеру вместе с кодовой таблицей.

Недостатками статического кода Хаффмана являются:

- дополнительное время, расходуемое на предварительный анализ текста;
- дополнительный объем информации, требуемый для передачи кодовой таблицы декодеру.

### Динамический код Хаффмана

Метод динамического кодирования Хаффмана реализован в ряде алгоритмов, в частности, FGK (Феллер, Галлагер, Кнут), предложенном в 1973 году.

Суть алгоритма – построение дерева Хаффмана с минимальным весом пути

$$\min \sum_{j=1}^m W_j l_j, \quad (2.16)$$

где

$W_j$  – вес символов  $z_j$ , поступивших на текущий момент времени;

$l_j$  – расстояние от корня до листа  $z_j$ ;

$m$  – мощность алфавита;

$z_j$  –  $j$ -й символ алфавита.

Также обозначим  $M(k) = z_1, z_2, \dots, z_k$  – первые  $k$  символов сообщения.

**Свойства оптимального дерева Хаффмана, имеющего  $m$  внешних узлов (листьев)**

1. Каждый лист имеет неотрицательный вес ( $W > 0$ ); каждый внутренний (родительский) узел имеет подчиненные (дочерние) узлы, а его вес равен сумме весов дочерних узлов.

2. На каждом уровне дерева (кроме корневого) должно быть не менее одной пары узлов, имеющей общий родительский узел.
3. Все узлы нумеруются в возрастающем порядке таким образом, что узлы с номером  $(2j-1)$  и  $2j$  являются узлами одного уровня ( $1 \leq j \leq m-1$ ), а их общий родительский узел имеет более высокий уровень.
4. Нумерация узлов соответствует тому порядку, в котором узлы объединяются в соответствии со статическим алгоритмом Хаффмана.

Процедура кодирования разбивается на два этапа:

1. Дерево Хаффмана после обработки  $M(k)$  преобразуется в эквивалентное.
2. Инкрементируется эквивалентное дерево.

Первый этап начинается после получения от источника символа  $z(k+1)$  с присвоением статуса текущего узла листу  $z(k+1)$ . Затем происходит обмен текущего узла (включая образованное им поддерево), с узлом, имеющим наибольший порядковый номер с таким же весом. После этого в качестве нового текущего узла инициализируется родительский узел последнего текущего узла. Обмен узлами может повторяться многократно, до тех пор, пока не будет достигнут корень. Максимальное количество перестановок может достигать высоты дерева  $l_{\max}$ .

На втором этапе инкрементируется лист дерева, соответствующий обрабатываемому символу и последующие промежуточные узлы, расположенные на пути движения от листа к корню дерева.

#### Пример 4

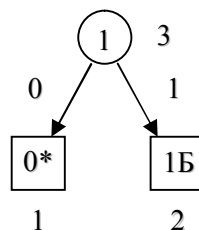
Требуется закодировать сообщение «БАРАБАН» с помощью динамического КХ. Если символ встречается первый раз, использовать ASCII коды.

Решение

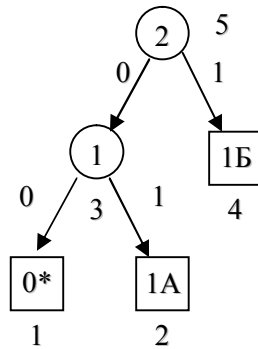
Обозначим исходное состояние кодера  $0^*$ .

Примечание. Если символ не встречается в обработанной последовательности, то лист  $0^*$  заменяется узлом ветвления, из которого выходят две ветви: левая – к узлу  $0^*$ , а правая – к символу  $z_{k+1}$ . Если кодер находится в исходном состоянии, то первый символ располагается вдоль правой ветви от корня.

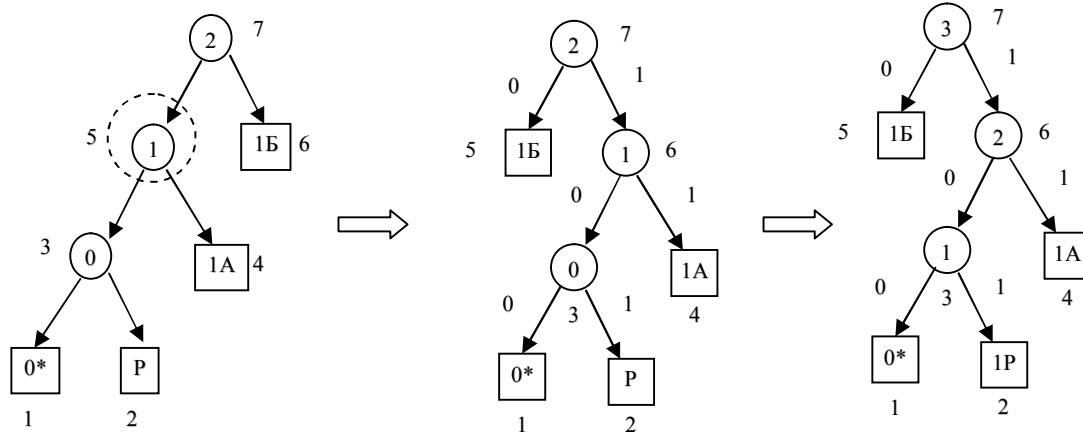
Шаг 1 – обработка символа Б (его ASCII-код отправляется в КС)



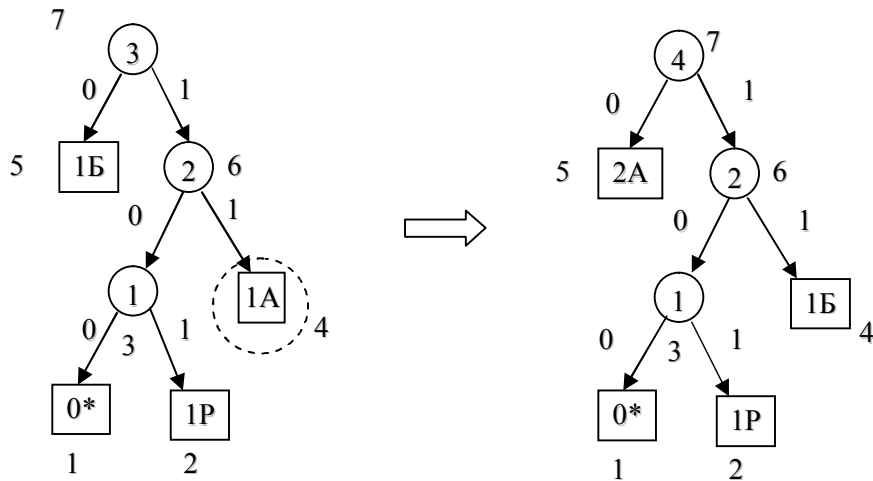
Шаг 2 – обработка символа А (его ASCII-код отправляется в КС)



Шаг 3 – обработка символа Р (его ASCII-код отправляется в КС)

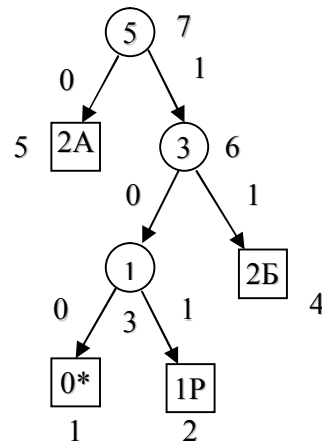
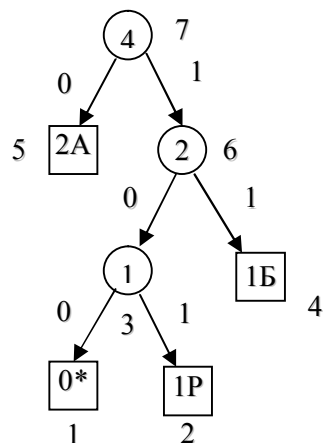


Шаг 4 – обработка символа А (код «11» его местоположения в дереве Хаффмана, построенном на третьем шаге, отправляется в КС)

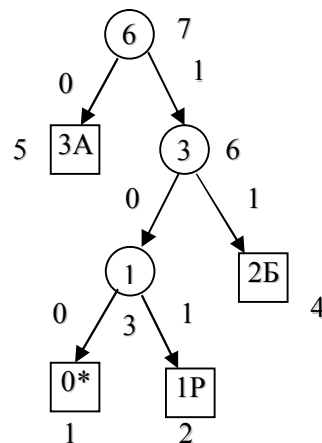
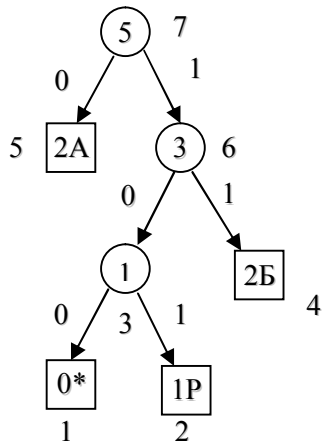




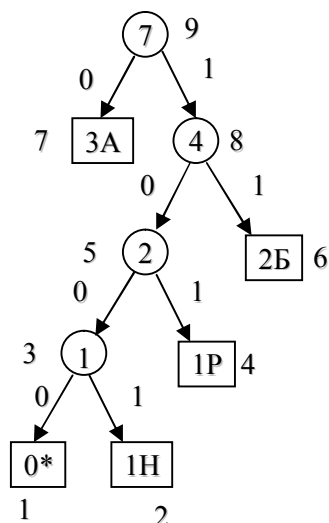
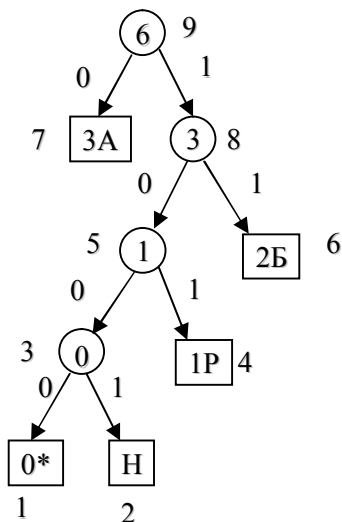
Шаг 5 – обработка символа Б (код «11» его местоположения в дереве Хаффмана, построенном на четвертом шаге, отправляется в КС)



Шаг 6 – обработка символа А (код «0» его местоположения в дереве Хаффмана, построенном на пятом шаге, отправлен в КС)



Шаг 7 – обработка символа Н (его ASCII-код отправляется в КС).



Расчет коэффициента сжатия слова «БАРАБАН» динамическим КХ  
 БАРАБАН – объем (V) несжатого сообщения 8 [бит] · 7=56 [бит];  
 БАР11110Н – объем (V<sub>сж</sub>) сжатого сообщения 8 [бит] · 4+5=37 [бит];  
 $K_{сж} = V_{сж} / V = 37/56 \approx 0,66$ .

### Арифметический код

При арифметическом кодировании (АК) текст представляется вещественными числами в интервале от 0 до 1. По мере кодирования текста отображающий его интервал уменьшается, а количество битов для его представления возрастает.

Очередные символы текста сокращают величину интервала исходя из значений их вероятностей, определяемых моделью источника. Более вероятные символы в меньшей степени сокращают интервал и, следовательно, добавляют меньше битов к результату, чем менее вероятные.

Перед началом работы тексту соответствует интервал [0; 1). При обработке очередного символа его ширина сужается за счет выделения этому символу части интервала.

### Пример 5

Требуется закодировать сообщение АВАВАСАД арифметическим кодом (АК).

Построение АК поясняют таблицы 2.4 и 2.5.

Таблица 2.4 – Статистическая таблица источника для построения АК

Символ	Вероятность	Интервал	НГ <sub>sk</sub>	ВГ <sub>sk</sub>
А	0,5	[0,0; 0,5)	0,0	0,5
В	0,25	[0,5; 0,75)	0,5	0,75
С	0,125	[0,75; 0,875)	0,75	0,875
Д	0,125	[0,875; 1,0)	0,875	1,0

В таблице приняты следующие обозначения:

НГ<sub>sk</sub> – нижняя граница k-го символа сообщения;

ВГ<sub>sk</sub> – верхняя граница k-го символа сообщения.

Для кодирования используют соотношения

$$\left. \begin{aligned} \text{НГ}_k &= \text{НГ}_{k-1} + I_k \times \text{НГ}_{sk} \\ \text{ВГ}_k &= \text{НГ}_{k-1} + I_k \times \text{ВГ}_{sk} \end{aligned} \right\} , \quad (2.17)$$

где  $I_k = (\text{ВГ}_{k-1} - \text{НГ}_{k-1})$  – k-й интервал ( $k = 1, 2, \dots, n$ );

n – количество символов в кодируемой последовательности;

ВГ<sub>k</sub>, НГ<sub>k</sub> – верхняя и нижняя границы на k-м шаге;

НГ<sub>k-1</sub> – нижняя границы на (k-1) - м шаге.

Результаты кодирования представлены в таблице 2.5

Таблица 2.5 - Результаты кодирования сообщения «ABABACAD»

Шаг алгоритма кодирования	Границы интервала
Начало	[0,0; 1,0)
После просмотра «А»	[0,0; 0,5)
После просмотра «В»	[0,25; 0,375)
После просмотра «А»	[0,25; 0,3125)
После просмотра «В»	[0,28125; 0,296875)
После просмотра «А»	[0,28125; 0,2890625)
После просмотра «С»	[0,287109375; 0,2880859375)
После просмотра «А»	[0,287109375; 0,28759765625)
После просмотра «D»	[0,28753662109375; 0,28759765625)

## Примечания

1. Кодеру и декодеру известен начальный интервал [0,1).
2. После просмотра первого символа («А») кодер сужает интервал до [0,0; 0,5), который модель выделяет этому символу.
3. Второй символ («В») сужает полученный на предыдущем шаге интервал до 1/4 части, задавая его середину в качестве нижней границы нового интервала, т.к. для «В» выделен фиксированный интервал [0,5; 0,75). В результате получается рабочий интервал [0,25; 0,375), т.к. предыдущий интервал имел длину 1/2 и  $1/2 \times 1/4 = 0,125$ .
4. Третий символ («А») сужает интервал, полученный на третьем шаге до [0,25; 0,3125), т.к. для «А» выделен фиксированный интервал [0,0; 0,5).
5. Четвертый символ («В») сужает интервал до [0,28125; 0,296875).
6. Пятый символ («А») сужает интервал до [0,28125; 0,2890625).
7. Шестой символ («С») сужает интервал до [0,287109375; 0,2880859375).
8. Седьмой символ («А») сужает интервал до [0,287109375; 0,28759765625).
9. Восьмой символ («D») сужает интервал до [0,28753662109375; 0,28759765625).

Учитывая, что результатом кодирования является любое число из конечного интервала, принимаем итоговое значение равным 0,28754.

Видно, что для кодирования сообщения из 8 символов потребовалось 5 десятичных чисел.

Данный вывод подтверждается расчетом энтропии источника в дитах

$$H = -4 \cdot \log 0,5 - 2 \cdot \log 0,25 - 2 \cdot \log 0,125 \approx 4,214 \approx 5 \text{ дит.}$$

Примечание. Значение энтропии можно получить, прологарифмировав величину конечного интервала кодирования:

$$H = -\log(0,28759765625 - 0,28753662109375) \approx -\log(0,000061) \approx 4,214 \approx 5 \text{ дит.}$$

На практике обычно применяют двоичное арифметическое кодирование.

**Пример 6**

Пусть дискретная случайная величина  $X$  (д.с.в.) принимает только два значения на выходе источника: 0 и 1 с вероятностями  $p_0 = 11/16$ ;  $p_1 = 5/16$  (см.таблицу 2.2). Сопоставим значению 0 отрезок [0,0; 11/16), а значению 1 — [11/16; 1,0).

Пусть длина блоков  $X$  равна трем:  $n=3$ . Значения интервалов, их вероятностей и кодов указаны в таблице 2.6.

Таблица 2.6 – Кодовая таблица двоичного АК ( $n=3$ )

Последовательность и интервал			Вероятность	Код
1→[11/16; 1,0)	11→[231/256, 1,0)	111→[3971/4096; 1,0)	125/4096	63/64=0,111111
		110→[3696/4096; 3971/4096)	275/4096	15/16= 0,1111
	10→[11/16; 231/256)	101→[3421/4096; 3696/4096)	275/4096	7/8= 0,111
		100→[2816/4096; 3421/4096)	605/4096	3/4= 0,11
0→[0,0; 11/16)	01→[121/256; 11/16)	011→[2541/4096; 2816/4096)	275/4096	5/8= 0,101
		010→[1936/4096; 2541/4096)	605/4096	1/2= 0,1
	00→[0,0; 121/256)	001→→[1331/4096; 1936/4096)	605/4096	3/8= 0,011
		000→[0,0; 1331/4096)	1331/4096	1/4= 0,01

### Алгоритм декодирования

1. По таблице 2.6 значений д.с.в. определяется интервал, содержащий текущий код. По этому интервалу однозначно определяется один символ исходного сообщения. Если этот символ — маркер конца сообщения, то конец.
2. Из текущего кода вычитается нижняя граница содержащего его интервала. Полученная разность делится на длину этого же интервала. Результат деления считается новым текущим значением кода. Переход к п. 1.

### Пример 7

1. Рассмотрим декодирование сообщения 111. Этому сообщению соответствует число  $7/8 \in [11/16; 1,0)$ . Т.е. первый знак декодируемого сообщения — 1.
2. Из  $7/8$  вычитается  $11/16$  и результат делится на  $5/16$ , что дает  $3/5 \in [0,0; 11/16)$ . Т.е. следующий знак — 0.
3. Вычисляем очередное текущее значение кода  $(3/5-0) \times 16/11=48/55 \in [11/16; 1,0)$ , т.е. следующий знак — 1.

Все исходное сообщение 101 декодировано.

Закодируем сообщение 0001 0001 0010 0011 из таблицы 2.2.

Таблица 2.7 - Результаты кодирования сообщения «0001 0001 0010 0011»

Шаг (i)	Символ	Нижняя граница	Верхняя граница	Ширина интервала
1	0	0	11/16	11/16
2	0	0	$121/16^2$	$121/16^2$
3	0	0	$1331/16^3$	$1331/16^3$
4	1	$14641/16^4$	$21296/16^4$	$6655/16^4$
5	0	$234256/16^5$	$307461/16^5$	$73205/16^5$

6	0	3748096/16 <sup>6</sup>	4553351/16 <sup>6</sup>	805255/16 <sup>6</sup>
7	0	59969536/16 <sup>7</sup>	68827341/16 <sup>7</sup>	8857805/16 <sup>7</sup>
8	1	1056948431/16 <sup>8</sup>	1101237456/16 <sup>8</sup>	44289025/16 <sup>8</sup>
9	0	17619799296/16 <sup>9</sup>	18106978571/16 <sup>9</sup>	487179275/16 <sup>9</sup>
10	0	281916788736/16 <sup>10</sup>	287275760761/16 <sup>10</sup>	5358972025/16 <sup>10</sup>
11	1	4569617312051/16 <sup>11</sup>	4596412172176/16 <sup>11</sup>	26794860125/16 <sup>11</sup>
12	0	73113876992816/16 <sup>12</sup>	73408620454191/16 <sup>12</sup>	294743461375/16 <sup>12</sup>
13	0	1169822031885056/16 <sup>13</sup>	1173064209960181/16 <sup>13</sup>	3242178075125/16 <sup>13</sup>
14	0	18717152510160896/16 <sup>14</sup>	18752816468987271/16 <sup>14</sup>	35663958826375/16 <sup>14</sup>
15	1	299866743709664461/16 <sup>15</sup>	300045063503796336/16 <sup>15</sup>	178319794131875/16 <sup>15</sup>
16	1	4799829417090082001/16 <sup>16</sup>	4800721016060741376/16 <sup>16</sup>	891598970659375/16 <sup>16</sup>

Нижней границе конечного интервала соответствует двоичное число 010000101001110001101011011011001001110000011111110010011010001<sub>2</sub> (или 429C.6B6C.9C1F.E4D116<sub>16</sub>).

Верхней границе конечного интервала соответствует двоичное число 0100001010011111100101100101010000101100000011011000011100000000<sub>2</sub> или 429F.9654.2C0D.8700<sub>16</sub>

В качестве результата кодирования принимаем сообщение с наименьшим числом значащих цифр, принадлежащее конечному интервалу: 429E.0000.0000.0000, деленное на 16<sup>16</sup>, т.е. 429E<sub>16</sub> → 0100.0010.1001.1110 или 010000101001111.

Видно, что после кодирования входного текста АК произошло сжатие сообщения с 16 до 15 бит.

### 3 ПОРЯДОК ВЫПОЛНЕНИЯ ИССЛЕДОВАНИЙ

- 3.1 Выбрать текстовое сообщение на естественном языке из таблицы А.1 в соответствии с вариантом задания.
- 3.2 Составить таблицу первичного алфавита источника на основе сообщения п.3.1 (модель Бернулли) с указанием частоты встречаемости символов.
- 3.3 Рассчитать информационные параметры источника и сообщения: количество информации, энтропию, избыточность.
- 3.4 Закодировать сообщение п.3.1 двоичным безизбыточным кодом.
- 3.5 Закодировать сообщение п.3.1 кодом Шеннона-Фано.
- 3.6 Закодировать сообщение п.3.1 статическим кодом Хаффмана.
- 3.7 Закодировать сообщение п.3.1 арифметическим кодом. Дополнительно, для первых восьми символов сообщения, записанного в п.3.4, выполнить двоичное кодирование и декодирование.
- 3.8 Рассчитать среднюю длину символа сообщения и определить коэффициенты сжатия для методов п.п.3.4-3.7.
- 3.9 Сделать выводы по работе.

### 4 СОДЕРЖАНИЕ ОТЧЕТА О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

- Титульный лист с указанием варианта задания;
- Название и цель работы;

- Теоретический раздел, в котором должны быть записаны все применяемые расчетные формулы;
- Результаты исследований по п.п. 3.1-3.9;
- Выводы, содержащие результаты анализа по п.п. 3.3-3.8.

## **5 КОНТРОЛЬНЫЕ ВОПРОСЫ**

- 5.1 Объясните значение информационного параметра дискретного источника «энтропия».
- 5.2 Запишите формулу для определения энтропии дискретного источника.
- 5.3 Какой смысл имеет энтропия с информационной точки зрения?
- 5.4 Как определяют количество информации дискретного источника?
- 5.5 В чем отличие количества и объема информации?
- 5.6 В каких случаях применяются формулы Шеннона и Хартли для определения количества информации?
- 5.7 Для чего применяют кодирование в схеме передачи информации?
- 5.8 Как определить избыточность текстового сообщения?
- 5.9 Назовите эффективные коды, изучаемые в данной лабораторной работе.
- 5.10 Дайте определение пропускной способности канала связи.
- 5.11 Какую пропускную способность должен иметь канал связи без помех, чтобы обеспечивать передачу информации от дискретного источника без задержек.
- 5.12 Назовите основное условие согласования источника информации и канала связи.
- 5.13 Как определяется скорость передачи информации в непрерывном канале связи?
- 5.14 Как определить границы, в которых находится средняя длина кодового слова?
- 5.15 Какое кодирование называется оптимальным (эффективным)?
- 5.16 Какими свойствами должны обладать оптимальные коды?
- 5.17 Приведите примеры равномерных и неравномерных эффективных кодов.
- 5.18 Объясните алгоритмы перевода кодового слова из двоичного безызбыточного кода в код Грея и обратно.
- 5.19 Почему код Грея называют зеркальным (или отраженным)?
- 5.20 Назовите основные особенности кода Грея.
- 5.21 Сформулируйте основные принципы оптимального кодирования.
- 5.22 Сформулируйте алгоритм Шеннона-Фано для двоичных алфавитов.
- 5.23 Сформулируйте алгоритм статического кода Хаффмана.
- 5.24 Объясните порядок кодирования сообщения динамическим кодом Хаффмана.
- 5.25 Назовите достоинства и недостатки статического и динамического кодов Хаффмана.
- 5.26 Объясните принцип арифметического кодирования.

### **Библиографический список**

1. Аветисян Р.Д., Аветисян Д.О. Теоретические основы информатики [Текст] / Р.Д. Аветисян, Д.О. Аветисян. — М. : Рос.гос.гуманит.ун-т, 1997. - 168 с.
2. Берлекэмп Э. Р. Алгебраическая теория кодирования [Текст] /Э.Р. Берлекемп. М.: Мир, 1971. — 478 с.
3. Дмитриев В.И. Прикладная теория информации [Текст] /В.И. Дмитриев. — М.: Высш. шк., 1989. — 320 с.
4. Кодирование информации (двоичные коды) [Текст]: справочник/Н.Т. Березюк, А.Г. Андрущенко, С.С. Мощицкий и др.; под ред. Н.Т. Березюка. — Харьков.: Вища шк., 1978. — 252 с.
5. Кузьмин И.В., Кедрус В.А. Основы теории информации и кодирования [Текст] /И.В. Кузьмин, В.А. Кедрус. К.: Виша шк., 1986. — 238 с.
6. Цымбал В.П. Задачник по теории информации и кодированию [Текст] / В.П. Цымбал — К.: Вища шк., 1992. — 276 с.
7. Цымбал В.П. Теория информации и кодирование [Текст]: учеб. для вузов / В.П. Цымбал — К.: Вища шк., 1992. — 263 с.

Приложение А  
(обязательное)  
Варианты заданий

1. Алтынного вора вешают, а полтинного чествуют.
2. Белый заяц бел, да цена ему пятнадцать копеек
3. Бог правду видит, да не скоро скажет.
4. Говорил попугай попугаю: "Я тебя, попугай, попугаю".
5. Вольно всякому на своей земле яму копать
6. Где нельзя перескочить, там можно перелезть.
7. Не жалела мама мыла, мама Милу мылом мыла.
8. Гость что чирей: где захочет, там и сядет.
9. Девушка Гагула села прясть да и заснула
10. Держись за авось, поколе не сорвалось
11. Носит Сеня в сени сено, спать на сене будет Сеня.
12. Если бы дрожать не умел, совсем бы замерз
13. За правду-матку ссылают в Камчатку
14. И строг наш приказ, да не слушают нас
15. Пришел Прокоп – кипит укроп, ушел Прокоп – кипит укроп.
16. Имеем - не храним, потеряем - плачем.
17. Какой бы муж не ворона, но жене оборона
18. Коза щиплет траву там, где ее привяжут
19. На дворе трава, на траве дрова, не руби дрова на траве двора.
20. Кто боязливо просит, тот учит отказывать
21. Кто малым недоволен, тот большого недостоин
22. Ласковый взгляд, да на сердце яд
23. Милости прошу к нашему шалашу: я пирогов покрошу и откусать попрошу.
24. Ленивый три раза ходит, а скупой три раза платит.
25. Лоб что лопата, а ума небогато.
26. Меня любишь, так и собаку мою не бей
27. Стоит поп на копне, колпак на попе, копна под попом, поп под колпаком.
28. Мир что вода: пошумит и разойдется
29. Мы люди не гордые: нету хлеба, подавай пироги
30. На золоте стоим, а нагнуться ленимся
31. Рапортовал, да не дорапортовал, а стал дорапортовывать, зарапортовался.
32. Не виновата курочка, что грязна улочка.
33. Не всегда вор крадет, а всегда берегись
34. Не долго думал, да ладно молвил
35. Был баран белокрыл, всех баранов перебелокрылил.
36. Не раскусив ореха, о ядре не толкуй
37. Не строй церкви, пристрой сироту
38. Не хвались умом, коли берешь все хребтом
39. Столы белодубовые гладкотёсовыструганные.
40. О чем не рассказывают, о том не допытывайся
41. Охал дядя, на чужие деньги глядя



42. Палка о двух концах: либо ты меня, либо я тебя
43. Белый снег. Белый мел. Белый сахар тоже бел.
44. Полдела около костра щепы подбирать.
45. Сколько кобылке не прыгать, а быть в хомуте.
46. Снявши голову, по волосам не плачут
47. Варвара караулила цыплят, а ворона воровала.
48. Три раза прости, а в четвертый прихворости
49. У корысти всегда рожа бескорыстна
50. У нашей Пелагеи все новые затеи
51. Борона боронила неборонованное поле.
52. У наших ворот всегда хоровод
53. У сотни безумных найдешь и умных
54. Удалось сидню с печи упасть
55. Бабкин боб расцвел в дождь, будет бабке боб в борщ.
56. Хорошо тому добро делать, кто помнит
57. Хоть конь горбат, да не мерину брат
58. Храни себя от бед, пока их нет
59. Бублик, баранку, батон и буханку пекарь из теста испек спозаранку.
60. Чем бы дитя не тешилось, лишь бы не плакало
61. Что в людях ведется, то и нас не минется
62. Что есть - вместе, а чего нет - пополам
63. Хлеб ржаной, батоны, булки не добудешь на прогулке.
64. Чужую бороду драть - своей не жалеть
65. Что у кого болит, тот о том и говорит
66. Хоть опоздать, да от людей не отстать
67. Еле-еле Лена ела, есть из лени не хотела.
68. Хоть денег ни гроша, да походка хороша
69. Хорошо тому скакать, кто не хочет пахать
70. Хозяин смекает, а гость примечает
71. Неделю Емеле прясть короб кудели, а Емелиной дочке - прясть одну ночку.
72. Хлеб да вода - крестьянская еда.
73. Аты-баты шли солдаты, аты-баты на базар. Аты-баты что купили? Аты-баты самовар!
74. Умный не осудит, а глупый не рассудит
75. У кого пороса пропало, тому и в ушах визжит.
76. Сколько не ело сито, ни разу не было сыто.
77. У наших бражников много праздников
78. Барашек на лугу гулял, барашек рожки потерял. Шел стороной голодный волк, барашка он зубами щелк!
79. Сшить не сошью, а распороть сумею
80. Сядем рядком да потолкуем ладком
81. Еле-еле Елизар, едет-едет на базар. А с базара, а с базара, не догонишь Елизара.
82. Собака помнит, кто ее кормит

83. Со своим уставом в чужой монастырь не ходят
84. Слово - серебро, молчание - золото.
85. Шел котик по лавочке, раздавал булабочки. Шел по скамеечке — раздавал копеечки.
86. Ругать жалко, а похвалить не за что
87. Рад бы заплакать, да смех одолел.
88. Птицы сильны крыльями, а люди дружбой.
89. Черепаха хвост поджала и за зайцем побежала. Оказалась впереди, кто не верит — выходи!
90. Правда хорошо, а счастье лучше
91. Уж как зоренька-заря русы косыньки плела. Кто те косы расплетет, первый кон водить идет!
92. Поменьше говори - больше услышишь
93. Кован-кован, перекован, у Ванюши конь подкован. Бьет подковой, бьет копытом, золотым гвоздем прибитым.
94. Подай палец, а за руку-то сам ухвачу
95. Катится яблочко с крутой горы. Кто поднимет — тот уйди!
96. По первому зову в гости не ездят
97. Ехала телега, сломалось колесо, сколько гвоздей на починку пошло?
98. Отойдем да поглядим: хорошо ли мы сидим
99. Вышел месяц из тумана, вынул пышку из кармана: — буду деточек кормить, а тебе, дружок, водить!
100. В первый день гость - золото, во второй день - серебро, а в третий день и гроша ломаного не стоит.