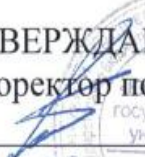


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 08.02.2021 17:26:09
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра космического приборостроения и систем связи

УТВЕРЖДАЮ
Проректор по учебной работе

О.Г. Локтионова
«22» 07 (ЮЗГУ) 2019 г.



ТЕОРИЯ ИНФОРМАЦИИ

Методические указания
по выполнению практических работ
для студентов, обучающихся по направлению подготовки*
10.03.01 «Информационная безопасность»
по дисциплине «Теория информации»

Курс 2019

УДК 621.391

Составители: Д.С. Коптев, И.Г. Бабанин

Рецензент:

Доктор технических наук, старший научный сотрудник,
профессор кафедры космического приборостроения и систем связи
В.Г. Андронов

Теория информации: методические указания по выполнению практических работ по дисциплине «Теория информации» / Юго-Зап. гос. ун-т; сост.: Д.С. Коптев, И.Г. Бабанин. – Курск, 2019. – 82 с.: иллюстр. 10., табл. 15. – Библиогр.: с. 82.

Методические указания по выполнению практических работ содержат краткие теоретические сведения об о природе информации, её свойствах, методах измерения её количества и качества, общих принципах кодирования информации в системах передачи, обработки и хранения, варианты заданий для выполнения работ, а также перечень вопросов для самопроверки изучаемого материала.

Методические указания соответствуют учебному плану по направлению подготовки 10.03.01 «Информационная безопасность», а также рабочей программе дисциплины: «Теория информации».

Предназначены для студентов, обучающихся по направлению подготовки 10.03.01 «Информационная безопасность», очной формы обучения.

Текст печатается в авторской редакции

Подписано в печать *22.07.19* Формат 60x84 1/16.
Усл. печ. л. 4,766. Уч.-изд. л. 4,315. Тираж 100 экз. Заказ *545* Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Практическая работа №1 «Этапы обращения информации»

1 Цель работы

Целью представленной лабораторной работы является ознакомление с этапами обращения информации и получение навыков сохранения, обработки и передачи данных при помощи различных технических средств.

2 Краткие теоретические сведения

2.1 Этапы обращения информации

Так как материальным носителем информации является сигнал, то можно следующим образом схематично представить этапы обращения и преобразования сигналов (рисунок 1).

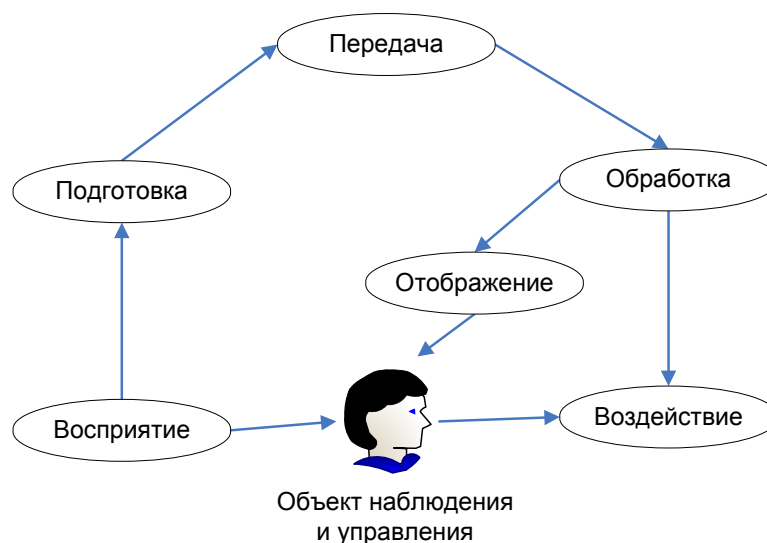


Рисунок 1 – Этапы обращения информации

Восприятие состоит в том, что формируется образ объекта, производится его опознание и оценка. При восприятии нужно отделить полезную информацию от шумов, что в некоторых случаях (тонкие биологические, физико-химические эксперименты; сложные производственные условия; радиосвязь, локация, астрономия и др.) связано со значительными трудностями. В результате восприятия получается сигнал в форме, удобной для передачи или обработки. В фазу восприятия могут включаться

операции подготовки информации, ее нормализации, квантования, кодирования, модуляции сигналов и построения моделей.

Передача информации состоит в переносе ее на расстояние посредством сигналов различной физической природы соответственно по механическим, гидравлическим, пневматическим, акустическим, оптическим, электрическим или электромагнитным каналам. Прием информации на другой стороне канала имеет характер вторичного восприятия со свойственными ему операциями борьбы с шумами.

Обработка информации заключается в решении задач, связанных с преобразованием информации, независимо от их функционального назначения. Применение ЭВМ обобщает и централизует функции обработки, имеющие отношение главным образом к моделям ситуаций и принятию решений при управлении. Обработка производится при помощи устройств или машин, осуществляющих аналоговые или цифровые преобразования поступающих величин и функций. Промежуточным этапом обработки может быть хранение в запоминающих устройствах. Извлечение информации из запоминающих устройств также имеет характер восприятия и связано с борьбой с помехами.

Представление информации требуется тогда, когда в цикле обращения информации принимает участие человек. Оно заключается в демонстрации перед человеком условных изображений, содержащих качественные и количественные характеристики выходной информации. Для этого используются устройства, способные воздействовать на органы чувств человека.

Из устройства обработки информация может выводиться не только оператору, но и непосредственно воздействовать на объект управления.

Воздействие состоит в том, что сигналы, несущие информацию, производят регулирующие или защитные действия, вызывая изменения в самом объекте.

2.2 Технологии сбора и хранения информации

Сбор предполагает получение максимально выверенной исходной информации и является одним из самых ответственных этапов в работе с информацией, поскольку от цели сбора и методов

последующей обработки полностью зависит конечный результат работы всей информационной системы.

Технология сбора подразумевает использование определенных методов сбора информации и технических средств, выбираемых в зависимости от вида информации и применяемых методов ее сбора. На заключительном этапе сбора, когда информация преобразуется в данные, т. е. в информацию, представленную в формализованном виде, пригодном для компьютерной обработки, осуществляется ее ввод в систему. Для сбора данных необходимо сначала определить технические средства, позволяющие осуществлять сбор быстро и высококачественно и поддерживающие операции ввода информации и представления данных в электронной форме.

На рисунке 2 схематично представлен процесс сбора информации.



Рисунок 2 - Процесс сбора информации

2.3 Технологический процесс обработки информации

Технологический процесс обработки информации — есть строго определенная последовательность взаимосвязанных процедур, выполняемых для преобразования первичной информации с момента ее возникновения до получения требуемого результата.

Технологический процесс призван автоматизировать обработку исходной информации за счет привлечения технических средств базовой информационной технологии, сократить финансовые и трудовые затраты, обеспечить высокую степень достоверности результатной информации. Для конкретной задачи той или иной предметной области технологический процесс обработки информации разрабатывается индивидуально. Совокупность процедур зависит от следующих факторов:

- характер и сложность решаемой задачи;
- алгоритм преобразования информации;
- используемые технические средства;
- сроки обработки данных;
- используемые системы контроля;
- число пользователей и т. д.

В общем случае технологический процесс обработки информации включает процедуры, представленный на рисунке 3.

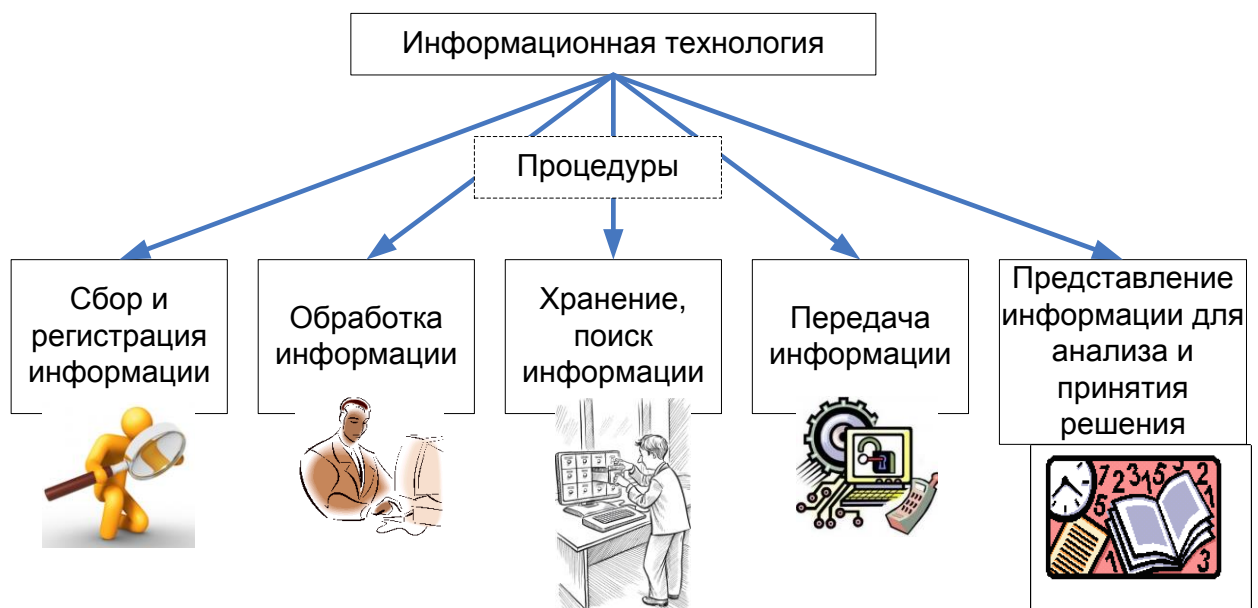


Рисунок 3 - Процедуры технологического процесса обработки информации

2.4 Способы обработки информации

Современные информационные технологии позволяют обрабатывать информацию централизованным и децентрализованным (т. е. распределенным) способами.

Централизованный способ предполагает сосредоточение данных в информационно-вычислительном центре, выполняющем все основные действия технологического процесса обработки информации. Основное достоинство централизованного способа — сравнительная дешевизна обработки больших объемов информации за счет повышения загрузки вычислительных средств.

Децентрализованный способ характеризуется рассредоточением информационно-вычислительных ресурсов и распределением технологического процесса обработки информации по местам возникновения и потребления информации. Достоинством децентрализованного способа является повышение оперативности обработки информации и решения поставленных задач за счет автоматизации деятельности на конкретных рабочих местах, применения надежных средств передачи информации, организации сбора первичных документов и ввода исходных данных в местах их возникновения [6].

Децентрализованный способ обработки информации может быть реализован автономным или сетевым методом. При автономной обработке информации передача документов и данных на электронных носителях осуществляется по почте либо курьером, а при сетевой — через современные каналы связи.

На практике применяют смешанный способ обработки информации, для которого характерны признаки двух способов одновременно (централизованный с частичной децентрализацией или децентрализованный с частичной централизацией).

2.5 Режимы обработки информации на компьютере

Вычислительные средства участвуют в процессе обработки информации в двух основных режимах: пакетном или диалоговом.

В случае, когда технология обработки информации на компьютере представляет собой заранее определенную последовательность операций, не требующую вмешательства человека, и диалог с пользователем отсутствует, информация обрабатывается в так называемом пакетном режиме. Суть его состоит в том, что программы обработки данных последовательно выполняются под управлением операционной системы как совокупность (пакет) заданий. Операционная система обеспечивает ввод данных, вызов требуемых программ, включение необходимых внешних устройств, координацию и управление технологическим процессом обработки информации.

Задачи, решаемые в пакетном режиме, характеризуются следующими свойствами:

- алгоритм решения задачи формализован, вмешательства пользователя не требуется;
- наличие большого объема входных и выходных данных, в основном хранящихся на устройствах хранения информации (например, жестких дисках компьютеров);
- расчет выполняется для большинства записей входных файлов;
- длительное время решения задачи — как правило, обусловлено большими объемами обрабатываемых данных;
- регламентность — задачи решаются с заданной периодичностью.

Пакетный режим возник первым и широко использовался с середины XX в., когда обработка информации на ЭВМ осуществлялась в специально создаваемых вычислительных центрах. Заказчики подготавливали исходные данные (обычно на перфокартах или перфолентах) и отправляли их в вычислительный центр, где данные обрабатывались и результаты обработки возвращались заказчику. С развитием персональных ЭВМ (начиная с 80-х гг. прошлого века) обработка данных стала осуществляться, в основном, непосредственно потребителями, поэтому в настоящее время пакетный режим используется достаточно редко. Сегодня более распространен диалоговый режим, когда необходимо непосредственное взаимодействие пользователя с компьютером и на каждое свое действие пользователь получает немедленные ответные действия компьютера. Диалоговый режим позволяет пользователю интерактивно управлять порядком обработки

информации и получать результатные данные в виде необходимых документов либо файлов.

2.6 Технологии передачи и представления информации

Информационные процессы невозможны без средств передачи и представления информации, поскольку зачастую информация требуется в месте, территориально удаленном от источника ее возникновения, и должна быть представлена в виде символов, образов и сигналов, пригодных для восприятия потребителем.

Современные средства связи способны передавать информацию в любой форме: телефонные, телевизионные, телеграфные сообщения, массивы данных, печатные материалы, фотографии и т. д. В соответствии со спецификой передаваемых сообщений организуется канал передачи информации — совокупность технических средств, обеспечивающих передачу сигналов от источника к потребителю.

Основная характеристика канала передачи — скорость передачи информации, а ее предельно допустимое значение называют емкостью канала, которая ограничивается шириной полосы канала и шумом.

Канал связи соединяет передатчик и приемник с помощью линии связи, которая может быть проводной, кабельной, радио, микроволновой, оптической или спутниковой. Примерами линий связи являются телефонные и вычислительные сети, сети телевизионного и радиовещания, мобильной связи, спутниковые технологии передачи данных.

В современных цифровых системах связи функции передатчика и приемника выполняет модем. Основное достоинство передачи информации в цифровой форме заключается в возможности использования кодированных сигналов, обеспечения защиты информации и наилучшего способа приема.

Для представления переданной или хранимой информации потребителю используются процессы воспроизведения и отображения.

Воспроизведение информации — это процесс, при котором ранее записанная на носителе информация считывается устройством воспроизведения.

Отображение информации — есть представление информации, т. е. генерация сигналов на основе исходных данных, а также правил и алгоритмов их преобразования в форме, приемлемой для непосредственного восприятия человеком.

Потребителем информации наиболее часто выступает человек, и для принятия решений ему необходимы результаты обработки информации. Тем не менее человек не способен ощутить машинное представление информации, а может воспринимать ее лишь органами чувств (органами зрения, слуха, осязания, обоняния и т. д.), поэтому для организации взаимодействия человека с информационными моделями объектов информационная система должна быть наделена специальными средствами отображения данных.

Поскольку зрение используется для восприятия информации наиболее активно, то средства отображения в современных ИС должны представлять информацию в лучшей форме для визуального наблюдения. Заметим, что мультимедиа-системы позволяют также представлять информацию в форме аудио- и видеосигналов, однако для управленческих информационных систем наиболее характерно отображение информации в текстовой и графической форме, что осуществляется за счет использования мониторов и печатающих устройств (например, принтеров, плоттеров).

Прежде чем, например, на мониторе, появится информация в доступной для человека форме, компьютером будет автоматически осуществлена следующая последовательность операций:

- преобразование данных, представленных в машинной форме, в вид, приемлемый для экранного отображения;
- согласование формы представления данных с параметрами монитора;
- воспроизведение в соответствии с возможностями воспроизводящего устройства (т. е. в данном примере — монитора).

3 Задание на практическую работу

В данной лабораторной работе предложено каждому студенту составить презентацию о себе и своем хобби.

Презентация должна быть выполнена при помощи программного продукта Microsoft PowerPoint версии 2007 и позднее.

Презентация должна содержать сведения об увлечениях студента, которые должны сопровождаться описанием увлечений, фотографиями, различными дипломами, грамотами и .т.д.

Слайды презентации должны меняться автоматически, быть красочными, яркими, привлекающими внимание.

3.1 Общие правила составления презентаций

Основное правило презентаций:

Простота, лаконичность (минимализм в подаче визуальной информации). Краткое изложение материала, максимальная информативность текста.

Следующие правила презентаций:

- Читабельность (видимость из самых дальних уголков помещения и с различных устройств);
- Отсутствие накопления, четкий порядок во всем.
- Тщательно структурированная информация.
- Наличие коротких и лаконичных заголовков, маркированных и нумерованных списков.
- Важную информацию (например, выводы, определения, правила и т.д.) нужно подавать большим и выделенным шрифтом и размещать в левом верхнем углу слайда.
- Второстепенную информацию желательно размещать внизу слайда.
- Каждому положению (идее) надо отвести отдельный абзац.
- Главную идею надо выложить в первой строке абзаца.
- Использовать табличные формы представления информации (диаграммы, схемы) для иллюстрации важнейших фактов, что даст возможность подать материал компактно и наглядно.
- Графика должна органично дополнять текст.

– Объяснение надо размещать как можно ближе к иллюстрациям, с которыми они должны появляться на экране одновременно.

– Инструкции к выполнению задач необходимо тщательно продумать относительно их четкости, лаконичности, однозначности.

– Использовать эмоциональный фон (художественная проза запоминается лучше, чем специальные тексты, а стихи — лучше, чем проза).

– Всю текстовую информацию нужно тщательно проверить на отсутствие орфографических, грамматических и стилистических ошибок.

– Производительность подаваемого материала увеличивается, если одновременно задействованы зрительный и слуховой каналы восприятия информации (зарубежные источники это называют принципом модальности). Поэтому рекомендуется там, где это возможно, использовать для текста и графических изображений звуковое сопровождение.

Исследования свидетельствуют, что эффективность слухового восприятия информации составляет 15 %, зрительного — 25 %, а их одновременное привлечение к процессу обучения повышает эффективность восприятия до 65 %.

3.2 Физиологические особенности восприятия цветов и форм

Стимулирующие (теплые) цвета способствуют возбуждению и действуют как раздражители (в порядке убывания интенсивности воздействия: красный, оранжевый, желтый).

Дезинтегрирующие (холодные) цвета успокаивают, вызывают сонливое состояние (в том же порядке: фиолетовый, синий, голубой, сине-зеленый, зеленый).

Нейтральные цвета: светло-розовый, желто-зеленый, коричневый.

Сочетание двух цветов — цвета знака и цвета фона — существенно влияет на зрительный комфорт, причем некоторые

пары цветов не только утомляют зрение, но и могут вызвать стресс (например: зеленые символы на красном фоне).

Лучшее сочетание цветов шрифта и фона: белый на темно-синем, черный на белом, желтый на синем, оранжевый на черном.

Цветовая схема должна быть одинаковой для всех слайдов.

Любой рисунок фона повышает утомляемость глаз и снижает эффективность восприятия информации.

Четкие, яркие рисунки, которые меняются, легко «охватывают» подсознание, и они лучше запоминаются.

Любой второстепенный объект, что движется (анимированный), снижает качество восприятия материала, отвлекает внимание, нарушает ее динамику.

Показ слайдов с фоновым сопровождением нежелательных звуков (песен, мелодий) вызывает быструю утомляемость, способствует рассеиванию внимания и снижает производительность обучения.

Человек может одновременно запоминать не более трех фактов, выводов, определений.

Каждый слайд должен отражать одну мысль.

Текст должен состоять из коротких слов и простых предложений.

Строка должна содержать 6-8 слов.

Всего на слайде должно быть 6-8 строк.

Общее количество слов не должно превышать 50.

Глаголы должны быть в одной временной форме.

Заголовки должны привлекать внимание аудитории и обобщать основные положения слайда.

В заголовках должны быть и большие, и малые буквы.

Слайды должны быть не слишком яркими — лишние украшения лишь создают барьер на пути эффективной передачи информации.

Количество блоков информации во время отображения статистических данных на одном слайде должно быть не более четырех.

Подписи к иллюстрации размещаются под ней, а не над ней.

Все слайды презентации должны быть выдержаны в одном стиле.

3.3 Общие правила использования шрифтов

1. Каждый шрифт (гарнитура¹ + написание) имеет одну смысловую нагрузку.

Для устойчивой гарнитуры традиционными, по меньшей мере, с XIX в. есть такие:

- полужирный шрифт названия структур документа,
- курсив — логическое ударение, в частности, на формулировании основных положений, определений и т.д.,
- «прямой» обычный - основной массив информации.

2. Тексты презентаций, которые используют в психологически напряженной нестандартной ситуации, надо подать гарнитурой с упрощенным алгоритмом распознавания, например, шрифтом Arial. Это целесообразно во время работы с инструкциями правил безопасности, нормативными актами, соглашениями с правовыми или имущественными последствиями, условиями олимпиадных заданий и т.п.

3. Избегайте использования более трех различных шрифтов на одном слайде. Иначе читатель преждевременно устанет, постоянно пытаясь выбрать алгоритм распознавания шрифта. Исключение составляет инструкция по использованию шрифтов.

4. Математические формулы представляются гарнитурой, близкой к стандартной (Times New Roman), причем все переменные — курсив, остальные — скобки, знаки математических действий, устоявшиеся названия функций (sin, cos и т.д.) - обычным «прямым» шрифтом.

3.5 Типичные недочеты и ошибки при создании презентаций

Типичными недочетами и ошибками при создании презентаций являются следующие:

- отсутствие Титульного слайда, содержащего: название проекта или темы урока (занятия); сведения об авторе; дата разработки; информация о местоположении ресурса в сети и др;

¹ **Гарнитура шрифта** - графический рисунок начертаний букв и знаков, составляющих единую стилистическую и композиционную систему.

- отсутствие Введения, в котором представлены: цели и задачи изучения темы, краткая характеристика содержания;
- отсутствие Оглавления (для развернутых разработок, при наличии в презентации разделов, подтем) с гиперссылками на разделы / подтемы презентации;
- отсутствие логического завершения презентации, содержащего: заключение, обобщения, выводы;
- перегрузка слайдов подробной текстовой информацией (не более трех мелких фактов на слайде и не более одного важного);
- неравномерное и нерациональное использование пространства на слайде;
- отсутствие связи фона презентации с содержанием.
- неудачный выбор цветовой гаммы: использование слишком ярких и утомительных цветов, использование в дизайне более 3 цветов(цвет текста, цвет фона, цвет заголовка и/или выделения); использование темного фона со светлым текстом;
- использование разных фонов на слайдах в рамках одной презентации;
- использование рисунков, фотографий плохого качества и с искажениями пропорций;
- отсутствие должного выравнивания текста;
- отсутствие или неясность связей в схемах или между компонентами материала на слайде;
- наличие различных эффектов переходов между слайдами и других раздражающих эффектов анимации, мешающих восприятию информации;
- отсутствие единства стиля страниц:
 - ✓ одинаковая гарнитура и размер шрифта для всех заголовков (не менее 24 пунктов);
 - ✓ одинаковая гарнитура и размер шрифта для тестовых фрагментов (не менее 18 пунктов);
 - ✓ заголовки, номера страниц, кнопки перелистывания должны появляться в одном и том же месте экрана;
 - ✓ одинаковая цветовая гамма на всех страницах и т.п.

4 Контрольные вопросы

1. Что такое сбор информации и каково его предназначение?
2. Что понимается под технологией сбора информации?
3. Чем отличаются понятия «информация» и «данные»?
4. Назовите основные требования к сбору данных и к хранимым данным.
5. Перечислите основные средства сбора текстовой, графической, звуковой и видеоинформации. Какие еще средства сбора информации вам известны?
6. Какие еще методы сбора данных вам известны?
7. В чем заключается процедура хранения информации?
8. Перечислите основные требования к структурам хранения.
9. Что такое база данных?
10. В чем различие между базой и банком данных?
11. Что такое резервное копирование и для чего оно осуществляется?
12. Что такое архивное копирование и в чем его отличие от резервного копирования?
13. Что такое базовая информационная технология?
14. В чем заключается различие между централизованным и децентрализованным способами обработки информации?
15. Какие режимы обработки информации вам известны?

Практическая работа №2 «Аддитивная мера информации (Мера Хартли)»

1 Цель работы

Изучить аддитивную меру информации. Научиться переводить числа из одной системы счисления в другую, научиться измерять количество информации для равновероятных и неравновероятных событий. Проверить полученные теоретические сведения практическим путем, выполняя индивидуальные задания.

2 Краткие теоретические сведения

2.1 Определение количества информации для равновероятных событий

Одно из направлений в измерении информации дает структурная теория, в которой количество информации определяется подсчетом информационных элементов или комбинаций из них.

Рассмотрим *аддитивную меру (меру Хартли)*. Из комбинаторики известно, что число сочетаний с повторениями из h элементов по l равно

$$Q = h^l \quad (1)$$

Таким образом, число всех двоичных кодовых комбинаций длины l равно 2^l ($h = 2$). В качестве меры информации Хартли предложил взять

$$H = \log_2 Q = \log_2 h^l = l. \quad (2)$$

Тогда 1 *бит* – это количество информации, содержащееся в двоичной кодовой комбинации единичной длины. Количество информации по Хартли эквивалентно количеству двоичных знаков "0" и "1" при кодировании сообщений по двоичной системе счисления.

Под системой счисления понимается способ представления любого числа с помощью некоторого алфавита символов, называемых цифрами.

2.2 Перевод числа из одной системы счисления в другую

Все системы счисления делятся на позиционные и непозиционные.

Непозиционными системами являются такие системы счисления, в которых каждый символ сохраняет свое значение независимо от места его положения в числе. Примером непозиционной системы счисления является римская система. К недостаткам таких систем относятся наличие большого количества знаков и сложность выполнения арифметических операций.

Система счисления называется позиционной, если одна и та же цифра имеет различное значение, определяющееся позицией цифры в последовательности цифр, изображающей число. Это значение меняется в однозначной зависимости от позиции, занимаемой цифрой, по некоторому закону. Примером позиционной системы счисления является десятичная система, используемая в повседневной жизни.

При переводе из любой системы счисления в десятичную, нужно целую и дробную части (если есть дробная) умножать по правилам перевода. При переводе числа из десятичной системы в другую, нужно целую часть делить, а дробную умножать по правилам перевода из одной системы счисления в другую.

2.3 Определение количества информации для событий с различными вероятностями.

Существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Рассмотрим примеры таких событий.

1. В коробке 20 карандашей, из них 15 красных и 5 чёрных. Вероятность вытащить наугад красный карандаш больше, чем чёрный.

2. При случайном падении бутерброда вероятность падения его маслом вниз (более тяжёлой стороной) больше, чем маслом вверх.

3. В пруду живут 8000 карасей, 2000 щук и 40000 пескарей. Самая большая вероятность для рыбака – поймать в этом пруду пескаря, на втором месте – карася, на третьем – щуку.

Количество информации в сообщении о некотором событии зависит от его вероятности. Чем меньше вероятность события, тем больше информации оно несёт.

$$p = K/N,$$

где K – количество случаев реализации одного из исходов события,
 N – общее число возможных исходов одного из событий

$$I = \log_2(1/p),$$

где I – количество информации,

p – вероятность события

Логарифмом числа a по основанию b называется показатель степени, в которую надо возвести число a , чтобы получить число b :

$$a^{\log_a b} = b, a > 0, b > 0, a \neq 1$$

Правило вычисления логарифмов чисел по основанию 2 с помощью электронного калькулятора:

$$\log_2 b = \log b / \log 2, \text{ где } \log b \text{ и } \log 2 \text{ – десятичные логарифмы.}$$

Порядок вычисления логарифма числа 6 по основанию 2 ($\log_2 6$) с помощью инженерного калькулятора:

$$6, \log, /, 2, \log, =$$

Количество информации в случае различных вероятностей событий определяется по формуле Шеннона:

$$I = -\sum_{i=1}^n p_i \cdot \log_2 p_i$$

где p_i – вероятность i -го события,

n – количество возможных событий

3 Примеры выполнения заданий

Пример 1

Рассмотрим систему, информационная емкость которой определяется десятичным числом $Q = 121$. Определим количество информации, содержащееся в системе, используя меру Хартли (2),

$$H = \log_2 Q = \log_2 121 \approx 6.91887 \approx 7 \text{ бит}$$

Заметим, что округление результата до целого необходимо проводить в сторону увеличения. Полученный результат означает, что при кодировании числа достаточно использовать 7 двоичных знаков (число возможных двоичных кодовых комбинаций равно $2^7 = 128$).

$$120 = 1 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 1111000_2.$$

Замечание. Разложение по двоичной системе производится для числа на 1 меньше в силу того, что отсчет ведется от нуля, а число комбинаций равно 121.

Для получения двоичного числа можно использовать метод последовательного деления числа на 2. При каждом делении определяется один двоичный знак кодовой комбинации: если деление без остатка – "0", в противном случае – "1". Рисунок 1 иллюстрирует данный метод. Столбец справа показывает двоичные знаки кодовой комбинации.

$$\begin{array}{r}
 120 \mid 2 \\
 \hline
 120 \mid 60 \mid 2 \\
 \hline
 0 \quad 60 \mid 30 \mid 2 \\
 \hline
 \quad 0 \quad 30 \mid 15 \mid 2 \\
 \hline
 \quad \quad 0 \quad 14 \mid 7 \mid 2 \\
 \hline
 \quad \quad \quad 1 \quad 6 \mid 3 \mid 2 \\
 \hline
 \quad \quad \quad \quad 1 \quad 2 \mid 1 \\
 \hline
 \quad \quad \quad \quad \quad 1
 \end{array}$$

Рисунок 1 - Перевод числа в двоичную систему счисления

Двоичное число выписывается в обратном порядке, двигаясь снизу вверх.

Пример 2

Перевести число 12201_3 из троичной системы счисления в десятичную.

Решение.

$$\begin{array}{cccccc}
 4 & 3 & 2 & 1 & 0 & \text{разряды} \\
 12201_3 = 1 \cdot 3^4 + 2 \cdot 3^3 + 2 \cdot 3^2 + 0 \cdot 3^1 + 1 \cdot 3^0 = 81 + 54 + 18 + 1 = 154_{10}
 \end{array}$$

Ответ: $12201_3 = 154_{10}$.

Пример 3

Перевести число $234,6_8$ из восьмеричной системы в десятичную.

Решение.

$$\begin{array}{cccc}
 2 & 1 & 0 & -1 \\
 234,6_8 = 2 \cdot 8^2 + 3 \cdot 8^1 + 4 \cdot 8^0 + 6 \cdot 8^{-1} = 128 + 24 + 4 + 0,75 = 156,75_{10}
 \end{array}$$

Ответ: $234,6_8 = 156,75_{10}$.

Пример 4

Перевести число $101,01_2$ из двоичной системы счисления в десятичную.

Решение:

2 1 0 -1 -2 разряды

$$101,01_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 0 \cdot 2^{-1} + 1 \cdot 2^{-2} = 4 + 0 + 1 + 0 + 0.25 = 5,25_{10}$$

Ответ: $101,01_2 = 5,25_{10}$.

Пример 5

Перевести дробное десятичное число $0,625_{10}$ в дробное двоичное число.

Решение.

$$0,625 \cdot 2 = 1,25 \quad 0,25 \cdot 2 = 0,5 \quad 0,5 \cdot 2 = 1$$

Ответ: $0,625_{10} = 0,101_2$.

Пример 6

Перевести число $0,165$ в четверичную систему счисления, ограничившись четырьмя четверичными разрядами..

Решение.

$$0,165 \cdot 4 = 0,66 \quad 0,66 \cdot 4 = 2,64 \quad 0,64 \cdot 4 = 2,56 \quad 0,56 \cdot 4 = 2,24$$

Ответ: $0,165_{10} = 0,0222_4$.

Пример 7

Перевести десятичное число 7467_{10} в шестнадцатеричную систему счисления.

Решение.

$$\begin{array}{r} 4767 \overline{)16} \\ 4756 \overline{)466} \overline{)16} \\ \quad 11 \overline{)464} \overline{)29} \overline{)16} \\ \quad \quad 2 \overline{)16} \overline{)1} \\ \quad \quad \quad 13 \end{array}$$

Ответ: $1D2B_{16}$.

Пример 8.

В коробке 32 карандаша, все карандаши разного цвета. Наугад вытащили красный. Какое количество информации при этом было получено?

Решение.

Так как вытаскивание карандаша любого цвета из имеющихся в коробке 32 карандашей является равновероятным, то число возможных событий равно 32.

$$N = 32, l = ?$$

$N = 2^l$, подставим значение $N = 32$, получим $32 = 2^5$, тогда $l = 5$ бит.

Ответ: 5 бит.

Пример 9.

В школьной библиотеке 16 стеллажей с книгами, на каждом – по 8 полок. Ученику сообщили, что нужный учебник находится на 2-ой полке 4-го стеллажа. Какое количество информации получил ученик?

Решение.

1) Число стеллажей (случаев) – 16, т.е. $N_1 = 16$,

Определим $N_1 = 2^{l_1}$, подставим числовое значение $16 = 2^4$, т.е. $16 = 2^4$, тогда $l_1 = 4$ бита.

2) Число полок на каждом стеллаже (случаев) – 8, т.е. $N_2 = 8$,

Определим $N_2 = 2^{l_2}$, подставим числовое значение $8 = 2^3$, т.е. $l_2 = 3$ бит.

3) $l = l_1 + l_2$, получим $l = 4$ бита + 3 бита = 7 бит.

Ответ: 7 бит.

Пример 10.

Загадывают число в диапазоне от 1 до 200. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет".

Решение.

Правильная стратегия состоит в том, чтобы количество вариантов каждый раз уменьшалось вдвое.

Например, загадано число 152.

1 вопрос: Число > 100? Да.

2 вопрос: Число < 150? Нет.

3 вопрос: Число > 175? Нет. и т.д.

Количество событий в каждом варианте будет одинаково, и их отгадывание равновероятно. $N = 2^l$, тогда $200 = 2^l$, т.е. $7 < l < 8$. Т.к. количество вопросов нецелым числом быть не может, то необходимо задать не более 8 вопросов.

Ответ: 8 вопросов

Пример 11

В коробке 50 шаров, из них 40 белых и 10 чёрных. Определить количество информации в сообщении о вытаскивании наугад белого шара и чёрного шара.

Решение.

Вероятность вытаскивания белого шара

$$p_1 = 40/50 = 0,8$$

Вероятность вытаскивания чёрного шара

$$p_2 = 10/50 = 0,2$$

Количество информации о вытаскивании белого шара

$$I_1 = \log_2(1/0,8) = \log_2 1,25 = \log 1,25 / \log 2 \approx 0,32 \text{ бит}$$

Количество информации о вытаскивании чёрного шара

$$I_2 = \log_2(1/0,2) = \log_2 5 = \log 5 / \log 2 \approx 2,32 \text{ бит}$$

Ответ: 0,32 бит, 2,32 бит

Пример 12.

В озере живут караси и окуни. Подсчитано, что карасей 1500, а окуней - 500. Сколько информации содержится в сообщениях о том, что рыбак поймал карася, окуня, поймал рыбу?

Решение.

События поимки карася или окуня не являются равновероятными, так как окуней в озере меньше, чем карасей.

Общее количество карасей и окуней в пруду

$$1500 + 500 = 2000.$$

Вероятность попадания на удочку карася

$$p_1 = 1500/2000 = 0,75,$$

Вероятность попадания на удочку окуня

$$p_2 = 500/2000 = 0,25.$$

$$I_1 = \log_2(1/p_1), I_2 = \log_2(1/p_2),$$

где I_1 и I_2 – вероятности поймать карася и окуня соответственно.

Количество информации в сообщении поймать карася и поймать окуня соответственно

$$I_1 = \log_2(1 / 0,75) \approx 0,43 \text{ бит}, I_2 = \log_2(1 / 0,25) \approx 2 \text{ бит}.$$

Количество информации в сообщении поймать рыбу (карася или окуня) рассчитывается по формуле Шеннона

$$I = - p_1 \cdot \log_2 p_1 - p_2 \cdot \log_2 p_2$$

$$I = - 0,75 \cdot \log_2 0,75 - 0,25 \cdot \log_2 0,25$$

$$I = -0,75 \cdot (\log 0,75 / \log 2) - 0,25 \cdot (\log 0,25 / \log 2)$$

$$I = 0,604 \text{ бит} \approx 0,6 \text{ бит.}$$

Ответ: в сообщении содержится 0,6 бит информации

4. Задания

Номер варианта определяется как последняя цифра номера зачетки студента. Студенту необходимо решить все задачи в соответствии с номером варианта.

Задание 1.

Перевести число из одной системы счисления в другую.

Вариант	Числа
1	$A_{10}=121$ в двоичную систему счисления
2	$A_2=10001010111,01$ в десятичную систему счисления
3	$A_{10}=135,656$ в двоичную систему счисления с точностью до пяти знаков после запятой
4	$A_8=345,766$ в двоичную систему счисления
5	$A_{10}=326$ в троичную систему счисления
6	$A_{10}=1211$ в пятеричную систему счисления
7	$A_{10}=0,625$ в двоичную систему счисления
8	$A_2=0,1101$ в десятичную систему счисления
9	$A_{10}=113$ в двоичную систему счисления
10	$A_{10}=96$ в троичную систему счисления

Задание 2.

Чему равно значение суммы чисел в восьмеричной системе счисления?

Вариант	Сумма
1	$1110101010_2 + 10111001_2$
2	$10111010_2 + 10010100_2$
3	$111101110,1011_2 + 1111011110,1_2$
4	$40F,4_{16} + 160,4_{16}$
5	$274,5_{16} + DD,4_{16}$
6	$120,3_{16} - B9,6_{16}$
7	$122_5 + 210_5$
8	$221_5 + 120_5$
9	$201_5 + 102_5$

10	$11_4 + 11_8 + 11_{16}$
----	-------------------------

Задание 3.

Определить количество информации (по Хартли), содержащееся в системе, информационная емкость которой характеризуется десятичным числом Q . Закодировать это число по двоичной системе счисления, до двух цифр после запятой.

вариант	1	2	3	4	5	6	7	8	9	10
Q	1350	850	1980	2150	1450	890	1300	970	1590	1700

Задание 4.

Вариант 1. Телеграфист за день работы принимает 64 сообщения. Все сообщения разные. Наугад выбрали одно сообщение. Какое количество информации при этом было получено?

Вариант 2. Ваш друг живет в 16-ти этажном доме. Сколько информации содержит сообщение о том, что друг живет на 7 этаже?

Вариант 3. Какое количество информации несет сообщение о том, что встреча назначена на июль?

Вариант 4. При угадывании целого числа в диапазоне от 1 до N было получено 8 бит информации. Чему равно N ?

Вариант 5. Защита курсовой работы назначена на 15-е число. Какое количество информации несет данное сообщение?

Вариант 6. В цеху находится 8 различных станков. Какое количество информации несет сообщение, что 1 станок вышел из строя?

Вариант 7. В корзине лежат 16 разноцветных шаров. Наугад вытащили белый шар. Какое количество информации при этом было получено?

Вариант 8. При угадывании целого числа в диапазоне от 1 до N было получено 4 бит информации. Чему равно N ?

Вариант 9. В доме вашего друга 8 подъездов. Сколько информации содержит сообщение о том, что друг живет в 5 подъезде?

Вариант 10. В столовой стоит 17 столиков. Сколько информации несет сообщение о том, что 1 столик свободен?

Задание 5.

Вариант 1. В кафедральной библиотеке 32 стеллажа с учебной литературой, в каждом – по 6 полок. Студенту сообщили, что нужный учебник находится на 4-ой полке 8-го стеллажа. Какое количество информации получил студент?

Вариант 2. Студенческую группу отправляют на практику на 4 недели по 6-ти предприятиям. Студенту сообщили, что на 3ей неделе практики он будет на 2м предприятии. Сколько информации содержит данное сообщение?

Вариант 3. День рождения студента приходится на 8 декабря. Сколько информации несет в себе данное сообщение?

Вариант 4. Какое количество информации несет сообщение о том, что встреча назначена на 25 января?

Вариант 5. В магазине радиодеталей 12 стеллажей по 6 полок на каждом. Сколько информации несет сообщение о том, что нужная деталь находится в 3ем стеллаже на 2й полке?

Вариант 6. Экскурсионная группа приехала в Санкт-Петербург на 5 дней. За это время группа должна посетить 12 музеев. Сколько информации несет сообщение о том, что в 3й день экскурсии группа посетит 4 музея?

Вариант 7. Защита курсовой работы назначена на 15-е мая. Какое количество информации несет данное сообщение?

Вариант 8. В 10-этажном доме вашего друга 6 подъездов. Сколько информации содержит сообщение о том, что друг живет в 5 подъезде на 3 этаже?

Вариант 9. В двух залах студенческой столовой стоит 24 столика. Сколько информации несет сообщение о том, что во втором зале 1 столик свободен?

Вариант 10. В аудитории 8 столов на 3х человек. Сколько информации содержится сообщении о том, что за 3м столом сидит 1 человек?

Задание 6.

Вариант 1. Загадывают число в диапазоне от 1 до 1000. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 2. Загадывают число в диапазоне от 1 до 75. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 3. Загадывают число в диапазоне от 1 до 500. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 4. Загадывают число в диапазоне от 1 до 50. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 5. Загадывают число в диапазоне от 1 до 750. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 6. Загадывают число в диапазоне от 1 до 150. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 7. Загадывают число в диапазоне от 1 до 10. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 8. Загадывают число в диапазоне от 1 до 120. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 9. Загадывают число в диапазоне от 1 до 800. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Вариант 10. Загадывают число в диапазоне от 1 до 250. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только "Да" или "Нет"

Задание 7.

Вариант 1. В корзине лежат 32 шара, среди них 4 белых, а остальные черные. Сколько битов информации содержится в сообщении о том, что из корзины вытащили белый шар?

Вариант 2. В алфавите некоторого языка всего две буквы. Каждое слово этого языка состоит из m букв. Известно, что можно составить 2048 различных слов. Сколько букв в каждом слове?

Вариант 3. В ящике 80 яблок, из них 55 красных и 25 зеленых. Определить количество информации в сообщении о вытаскивании наугад красного и зеленого яблок.

Вариант 4. В книжном шкафу находятся 129 методических указаний, из них 45 по Цифровым системам передачи, остальные - по Общей теории связи. Сколько информации содержит сообщение о том, что наугад вытасченное методическое указание будет по Общей теории связи?

Вариант 5. Определите, сколько бит информации несет сообщение о том, что на светофоре горит зеленый свет

Вариант 6. Шахматная доска состоит из 64 полей: пополам белых и черных. Сколько информации несет сообщение о том, что белая королева стоит на черной клетке?

Вариант 7. Предположим, вероятность того, что вы получите за контрольную работу оценку "5", равна 0,6; вероятность получения "4" равна 0,2; вероятность получения "3" - 0,2. Определите, сколько бит информации будет нести сообщение об оценке "3"?

Вариант 8. Предположим, вероятность того, что вы получите за контрольную работу оценку "5", равна 0,6; вероятность получения "4" равна 0,2; вероятность получения "3" - 0,2. Определите, сколько бит информации будет нести сообщение об оценке "4"?

Вариант 9. Предположим, вероятность того, что вы получите за контрольную работу оценку "5", равна 0,6; вероятность получения "4" равна 0,2; вероятность получения "3" - 0,2. Определите, сколько бит информации будет нести сообщение об оценке "5"?

Вариант 10. Сообщение занимает 3 страницы по 25 строк. Сколько информации содержит сообщение о том, что искомое слово находится на 2й странице в 17 строке?

Задание 8.

Вариант 1. Книга, набранная с использованием текстового редактора, содержит 70 страниц, на каждой странице 38 строк, в каждой строке 56 символов. Определить объем информации, содержащейся в книге.

Вариант 2. Измерьте информационный объем сообщения "Ура! Скоро Новый год!" в битах, байтах, килобайтах (Кб), мегабайтах (Мб).

Указание: считается, что текст набран с помощью компьютера, один символ алфавита несет 1 байт информации. Пробел – это тоже символ в алфавите мощностью 256 символов.

Вариант 3. Измерьте примерную информационную емкость одной страницы любого своего учебника, всего учебника.

Указание: Для выполнения задания возьмите учебник по любимому предмету, посчитайте число строк на странице, число символов в строке, включая пробелы. Помните, что один символ алфавита несет 1 байт информации. Перемножив полученные значения, Вы найдете информационную емкость одной страницы учебника (в байтах).

Вариант 4. Проводят две лотереи: «4 из 32» и «5 из 64». Сообщение о результатах какой из лотерей несет больше информации?

Вариант 5. Информационное сообщение объемом 1.5 Кбайта содержит 3072 символа. Сколько символов содержит алфавит, при помощи которого было записано это сообщение?

Вариант 6. Книга, набранная с помощью компьютера, содержит 150 страниц; на каждой странице — 40 строк, в каждой строке — 60 символов. Каков объем информации в книге?

Вариант 7. Подсчитайте объем информации, содержащейся в романе А. Дюма "Три мушкетера", и определите, сколько близких по объему произведений можно разместить на одном лазерном диске? (590 стр., 48 строк на одной странице, 53 символа в строке).

Вариант 8. На диске объемом 100 Мбайт подготовлена к выдаче на экран дисплея информация: 24 строчки по 80 символов, эта информация заполняет экран целиком. Какую часть диска она занимает?

Вариант 9. Сообщение занимает 3 страницы по 25 строк. В каждой строке записано по 60 символов. Сколько символов в использованном алфавите, если все сообщение содержит 1125 байтов?

Вариант 10. Измерьте примерную информационную емкость одной страницы учебника, всего учебника. Сколько таких учебников может поместиться на дискете емкостью 360 Кбайт, 1.44 Мбайт, на винчестере в 420 Мбайт, в 6,4 Гбайт ?

5 Контрольные вопросы

1. Как определяется количество информации по Хартли?
2. Дайте определение количества информации равной 1 биту.
3. Что понимается под системой счисления?
4. Какие системы счисления являются непозиционными?
5. Дайте определение позиционным системам счисления.
6. Приведите алгоритм перевода числа из десятичной в любую другую систему счисления
7. В чем заключается алгоритм перевода из любой системы счисления в десятичную?
8. Какие события называют равновероятными?
9. Какие события являются неравновероятными?
10. Как определить количество информации для события с равновероятными символами?
11. Как определить количество информации для события, символы которого неравновероятны?
12. В каких случаях сообщение несет максимальное количество информации?
13. Когда количество информации равно нулю?

Практическая работа №3 «Расчет количества информации и энтропии для дискретных сообщений»

1 Цель работы

- приобрести умение определять количество информации и энтропию для дискретного источника сообщений.

2 Основные теоретические сведения

Источник сообщений, который может в каждый момент времени случайным образом принять одно из конечного множества возможных состояний, называют **дискретным источником сообщений**. Каждому состоянию источника и ставится в соответствие условное обозначение в виде символа из алфавита этого источника: S_1, S_2, \dots, S_m .

Источник сообщений в общем случае характеризуется ансамблем S , то есть полной совокупностью состояний с вероятностями их появления, составляющими в сумме единицу:

$$S = \left(\begin{array}{cccc} s_1 & s_2 & \dots & s_m \\ p(s_1) & p(s_2) & \dots & p(s_m) \end{array} \right); \text{ причём } \sum_{i=1}^m p(s_i) = 1$$

Энтропия $H(S)$ дискретного источника сообщений, задаваемого ансамблем S , может быть определена по формуле Шеннона:

$$H(S) = - \sum_{i=1}^m p(s_i) \log_2 p(s_i) \text{ (бит/символ);}$$

где $p(s_j)$ - вероятность появления символа s_j из алфавита объемом m .

Количество информации в сообщении, состоящем из n символов, может быть найдено следующим образом:

$$I = n \cdot H(S) = -n \sum_{i=1}^m p(s_i) \log_2 p(s_i) \text{ (бит)}$$

При наличии взаимозависимости символов алфавита используется **условная энтропия**. С учетом статистической связи между парами символов условная энтропия может быть найдена из формулы:

$$H_2(S) = - \sum_{i=1}^m \sum_{j=1}^m p(s_i) p(s_j | s_i) \log_2 p(s_j | s_i) \text{ (бит/символ);}$$

где $p(S_j | s_i)$ - условная вероятность появления символа S_j при заданном символе S_i .

Другой важной характеристикой источника сообщений, помимо энтропии, является **информационная избыточность**, которая показывает относительную недогруженность на символ алфавита источника сообщений:

$$R = \frac{H_{\max} - H}{H_{\max}} = 1 - \frac{H}{H_{\max}};$$

где $H_{\max} = \log_2 m$ - максимально возможная для данного источника сообщений энтропия, которая достигается при равной вероятности появления символов.

Кроме общей информационной избыточности существуют частные избыточности, основными из которых являются:

1) Избыточность, вызванная статистической связью между символами s_1, s_2, \dots, s_m :

$$R_p = 1 - \frac{H}{H'};$$

где $H' = - \sum_{i=1}^m p(s_i) \log_2 p(s_i)$ - энтропия источника сообщений при отсутствии статистической связи между символами.

2) Избыточность, вызванная не экстремальным распределением символов s_1, s_2, s_m (не равновероятным появлением символов в сообщениях):

$$R_\varphi = 1 - \frac{H'}{H_{\max}}.$$

3) Общая информационная избыточность связана с частными избыточностями следующим образом:

$$R = R_p + R_\varphi - R_p R_\varphi.$$

3 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.

3) Определить энтропию дискретного источника сообщений, алфавит которого состоит из восьми независимых символов S_1, S_2, \dots, S_8 . Известны вероятности появления символов $p(s_1)=p_1, p(s_2)=p_2, \dots, p(s_8)=p_8$ (таблица 1).

4) Определить среднее и частное количество информации, содержащейся в сообщении, которое состоит из ваших фамилии, имени и отчества (между словами сообщения допускаются пробелы). Вероятности появления букв русского алфавита взять из табл. П.1. Сравнить полученные результаты.

5) В условии задачи, сформулированной в пункте 3, учесть зависимость между парами символов, которая задана матрицей условных вероятностей $p(s_j|s_i)$ (табл. 2 – 5).

6) Для задачи, сформулированной в пункте 4, учесть наличие статистической связи между парами и тройками букв.

7) Определить частные избыточности, вызванные наличием связи между символами и не экстремальным распределением символов, для полученных при выполнении пунктов 5 и 6 результатов. На основе частных избыточностей найти общую информационную избыточность источника сообщений.

8) Оформить и защитить отчет по лабораторной работе.

4 Варианты заданий

Номер варианта студента определяется как порядковый номер студента в журнале преподавателя.

Таблица 1 – Вероятности появления независимых символов s_1, s_2, \dots, s_8

№ вар.	Вероятности для символов s_1, s_2, \dots, s_8							
	P1	P2	P3	P4	P5	P6	P7	P8
1	0,21	0,12	0,05	0,08	0,03	0,14	0,2	0,17
2	0,13	0,04	0,19	0,16	0,18	0,12	0,13	0,05
3	0,2	0,15	0,16	0,04	0,19	0,05	0,12	0,09
4	0,14	0,1	0,04	0,17	0,12	0,16	0,05	0,22
5	0,06	0,21	0,03	0,23	0,13	0,12	0,14	0,08
6	0,08	0,19	0,22	0,2	0,09	0,03	0,17	0,02
7	0,09	0,18	0,06	0,12	0,15	0,1	0,07	0,23
8	0,22	0,13	0,12	0,15	0,07	0,04	0,19	0,08
9	0,1	0,24	0,08	0,07	0,17	0,05	0,16	0,13

10	0,04	0,17	0,15	0,11	0,1	0,09	0,13	0,21
11	0,17	0,05	0,16	0,22	0,08	0,19	0,09	0,04
12	0,12	0,09	0,08	0,14	0,16	0,21	0,04	0,16
13	0,23	0,1	0,07	0,06	0,05	0,12	0,22	0,15
14	0,16	0,08	0,05	0,18	0,11	0,15	0,07	0,2
15	0,24	0,12	0,13	0,09	0,14	0,08	0,15	0,05
16	0,11	0,04	0,19	0,16	0,18	0,12	0,13	0,07
17	0,09	0,23	0,05	0,21	0,15	0,1	0,12	0,05
18	0,2	0,15	0,1	0,14	0,08	0,06	0,17	0,1
19	0,15	0,06	0,16	0,23	0,07	0,19	0,08	0,06
20	0,22	0,12	0,06	0,07	0,04	0,13	0,2	0,16
21	0,13	0,08	0,1	0,09	0,21	0,19	0,05	0,15
22	0,08	0,2	0,07	0,13	0,14	0,09	0,05	0,24

Таблица 2 – Матрица условных вероятностей для вариантов 1, 5, 9, 13, 17, 21

i,j	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	0,14	0,22	0	0,27	0,09	0	0,12	0,16
S_2	0,25	0	0,13	0,08	0,2	0,13	0,21	0
S_3	0	0,26	0,16	0,22	0,12	0,09	0	0,15
S_4	0,07	0,18	0,24	0	0,29	0	0,1	0,12
S_5	0,09	0,3	0,08	0,21	0	0,17	0	0,15
S_6	0	0,14	0,15	0	0,28	0,07	0,23	0,13
S_7	0,15	0,06	0	0,27	0,24	0,08	0,2	0
S_8	0,23	0	0,16	0,1	0	0,19	0,07	0,25

Таблица 3 – Матрица условных вероятностей для вариантов 2, 6, 10, 14, 18, 22

h_i	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	0	0,17	0,28	0	0,13	0,08	0,23	0,11
S_2	0,13	0,05	0	0,26	0,19	0,13	0	0,24
S_3	0,14	0	0,1	0,15	0,29	0	0,07	0,25
S_4	0,07	0,18	0,24	0	0,28	0	0,1	0,13
S_5	0,26	0	0,19	0,08	0	0,1	0,28	0,09
S_6	0,06	0,24	0	0,17	0,09	0,33	0	0,11
S_7	0	0,08	0,15	0,27	0,22	0,08	0,2	0
S_8	0,09	0,25	0,14	0,16	0	0,21	0,15	0

Таблица 4 – Матрица условных вероятностей для вариантов 3, 7, 11, 15, 19, 23

i,j	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	0,24	0,2	0	0,15	0,22	0	0,14	0,05
S_2	0,15	0	0,08	0,14	0,19	0,23	0,21	0
S_3	0	0,14	0,27	0,24	0,15	0,11	0	0,09
S_4	0,09	0,13	0,28	0	0,25	0	0,15	0,1
S_5	0,11	0,22	0,2	0,07	0	0,27	0	0,13
S_6	0	0,06	0,09	0	0,2	0,15	0,29	0,21
S_7	0,1	0,25	0	0,08	0,14	0,17	0,26	0
S_8	0,06	0	0,3	0,14	0	0,23	0,07	0,2

Таблица 5 – Матрица условных вероятностей для вариантов 4, 8, 12, 16, 20, 24

i,j	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	0	0,18	0,07	0	0,3	0,12	0,13	0,2
S_2	0,29	0,11	0	0,08	0,21	0,17	0	0,14
S_3	0,06	0	0,15	0,26	0,18	0	0,07	0,28
S_4	0,14	0,09	0,3	0	0,22	0	0,19	0,06
S_5	0,2	0	0,27	0,07	0	0,1	0,21	0,15
S_6	0,15	0,21	0	0,06	0,11	0,29	0	0,18
S_7	0	0,12	0,28	0,15	0,09	0,2	0,16	0
S_8	0,1	0,07	0,18	0,29	0	0,21	0,15	0

5. Контрольные вопросы

- 1) Что называют дискретным источником сообщений?
- 2) Что характеризует энтропия сообщения, и каким образом она может быть определена для дискретного источника сообщений?
- 3) В каком случае энтропия дискретного источника сообщений принимает максимальное значение?
- 4) Как можно найти количество информации, содержащееся в дискретном сообщении?
- 5) Как определяется условная энтропия источника сообщений при наличии статистической связи между состояниями этого источника?

6) Что характеризует информационная избыточность источника сообщений?

7) Какие выделяют частные виды информационной избыточности и как они связаны с общей избыточностью источника сообщений?

Практическая работа №4

«Расчёт условной энтропии дискретных сообщений, передаваемых по каналу связи с помехами»

1 Цель работы

- приобрести умение определять условную энтропию, энтропию объединения и информационные потери для дискретных сообщений, передаваемых по каналу связи, в котором присутствуют помехи.

2 Основные теоретические сведения

Понятие условной энтропии в теории информации используется при определении взаимозависимости между символами кодируемого алфавита, для определения потерь при передаче информации по каналам связи, при вычислении энтропии объединения.

При нахождении условной энтропии в том или ином виде используются условные вероятности.

Если при передаче N сообщений символ a появится M раз, символ b появится L раз, а символ a вместе с символом b – K раз, то вероятности появления символов a и b будут $p(a)=M/N$ и $p(b)=L/N$; вероятность совместного появления символов a и b $p(a,b)=K/N$. Отсюда условная вероятность появления символа b относительно символа a и условная вероятность появления символа a относительно символа b будут:

$$p(b|a)=p(a,b)/p(a)=(K/N)/(M/N)=K/M;$$

$$p(a|b)=p(a,b)/p(b)=(K/N)/(L/N)=K/L.$$

При известной условной вероятности можно определить вероятности совместного появления символов:

$$p(a,b)=p(a)p(b|a)=p(b)p(a|b).$$

Условная энтропия источника сообщений B относительно источника A характеризует степень неопределенности в сообщении из источника B , при известном сообщении из источника A .

Условная энтропия дискретного источника сообщений B

относительно источника сообщений A может быть определена следующим образом:

$$H(B|A) = -\sum_{i=1}^m \sum_{j=1}^m p(a_i) p(b_j|a_i) \log_2 p(b_j|a_i) = -\sum_{i=1}^m \sum_{j=1}^m p(a_i, b_j) \log_2 p(b_j|a_i).$$

В общем случае, если передается m символов источника A и ожидается получить m символов источника B , то влияние помех в канале связи полностью описывается **канальной матрицей**:

h_j	b_1	b_2	...	b_j	...	b_m
a_1	$p(a_1 b_1)$	$p(a_1 b_2)$...	$p(a_1 b_j)$...	$p(a_1 b_m)$
a_2	$p(a_2 b_1)$	$p(a_2 b_2)$...	$p(a_2 b_j)$...	$p(a_2 b_m)$
...	, • •	, • •
a_i	$p(a_i b_1)$	$p(a_i b_2)$...	$p(a_i b_j)$...	$p(a_i b_m)$
... • •
a_m	$p(a_m b_1)$	$p(a_m b_2)$...	$p(a_m b_j)$...	$p(a_m b_m)$

Вероятности, расположенные по диагонали (выделены полужирным шрифтом) канальной матрицы, определяют правильный прием, остальные - ложный. Вероятности, заполняющие столбцы канальной матрицы, обычно уменьшаются по мере удаления от главной диагонали и при полном отсутствии помех все, кроме вероятностей, расположенных на главной диагонали, равны нулю.

Энтропия объединения $H(A,B)$ характеризует неопределенность того, что было послано сообщение из источника A , а принято сообщение из источника B . Для ансамблей переданных сообщений A и принятых сообщений B энтропия объединения представляет собой сумму вида:

$$H(A,B) = -\sum_{i=1}^m \sum_{j=1}^m p(a_i, b_j) \log_2 p(a_i, b_j) \text{ (бит/два символа);}$$

где $p(a_i, b_j)$ - вероятность совместного появления в сообщениях символов a_i и b_j из ансамблей A и B .

Энтропия объединения и условная энтропия связаны между собой соотношениями:

$$H(A,B) = H(A) + H(B|A) = H(B) + H(A|B);$$

$$H(B|A) = H(A,B) - H(A); \quad H(A|B) = H(A,B) - H(B).$$

Энтропия объединения может быть найдена на основе матрицы совместных вероятностей:

u	b_1	b_2	\dots	b_m
d_1	$p(a_1, b_1)$	$p(a_1, b_2)$	\dots	$P(a_1, b_m)$
d_2	$p(a_2, b_1)$	$p(a_2, b_2)$	\dots	$p(a_2, b_m)$
\dots	\dots	\dots	\dots	\dots
d_m	$P(a_m, b_1)$	$p(a_m, b_2)$	\dots	$P(a_m, b_m)$

Такая матрица обладает следующим свойством:

$$\sum_{i=1}^m p(a_i, b_i) = p(b_i); \quad \sum_{j=1}^m p(a_i, b_j) = p(a_i); \quad \text{при этом} \quad \sum_{i=1}^m p(a_i) = \sum_{j=1}^m p(b_j) = 1.$$

Данное свойство, в свою очередь, позволяет вычислить энтропию, как источника, так и приемника сообщений непосредственно по канальной матрице:

$$H(A) = -\sum_{i=1}^m \sum_{j=1}^m p(a_i, b_j) \log_2 \sum_{j=1}^m p(a_i, b_j);$$

$$H(B) = -\sum_{i=1}^m \sum_{j=1}^m p(a_i, b_j) \log_2 \sum_{j=1}^m p(a_i, b_j);$$

На основе условной энтропии и энтропии объединения можно определить потери информации в канале связи с шумами. **Информационные потери** при передаче n символов по данному каналу связи можно найти из формулы:

$$\Delta I = n \cdot H(A|B)$$

Отсюда среднее количество принятой информации может быть определено следующим образом:

$$I = n [H(B) - H(A|B)] = n \cdot H(B) - \Delta I.$$

3 Порядок выполнения работы

Данная лабораторная работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) Составить канальную матрицу, описывающую данный канал связи с точки зрения условий прохождения символов s_1, s_2, s_3, s_4 , для которых известны частоты появления в сообщении (табл. 2.1).
- 4) Определить общую условную энтропию сообщений,

алфавитом которых ЯВЛЯЮТСЯ СИМВОЛЫ S_1, S_2, S_3, S_4 , если известны вероятности появления этих символов в передаваемых сообщениях (табл. 2.2).

- 5) Вычислить энтропию принятых сообщений.
- 6) Составить матрицу совместных вероятностей.
- 7) Определить энтропию объединения.
- 8) Найти информационные потери при передаче 500 символов алфавита s_1, s_2, s_3, s_4 и количество принятой информации.
- 9) Оформить и защитить отчет по практической работе.

4 Варианты заданий

Номер варианта определяется как порядковый номер студента в журнале преподавателя.

Таблица 1 – Частоты появления символов на выходе канала связи при 150 случаях передачи каждого символа

Вар.	Пер. симв.	Принятые символы				Вар.	Пер. симв.	Принятые символы			
		s_1	s_2	s_3	s_4			s_1	s_2	s_3	s_4
1	S_1	148	0	2	0	4	S_1	146	1	2	1
7	S_2	0	147	1	2	10	S_2	0	148	1	1
13	S_3	1	0	149	0	16	S_3	0	3	147	0
19	S_4	0	3	1	146	22	S_4	1	0	0	149
2	S_1	147	2	0	1	5	S_1	147	2	1	0
8	S_2	1	149	0	0	11	S_2	3	146	0	1
14	S_3	0	2	146	2	17	S_3	0	1	149	0
20	S_4	2	0	0	148	23	S_4	0	2	0	148
3	S_1	149	1	0	0	6	S_1	148	1	1	0
9	S_2	1	146	3	0	12	S_2	2	147	1	0
15	S_3	2	0	148	0	18	S_3	2	1	146	1
21	S_4	0	1	2	147	24	S_4	0	0	1	149

Таблица 2 – Вероятности появления символов в сообщениях

Вар.	Символы алфавита				Вари- ант	Символы алфавита			
	s_1	s_2	s_3	s_4		s_1	s_2	s_3	s_4
1	0,38	0,21	0,32	0,09	13	0,09	0,15	0,41	0,35
2	0,27	0,39	0,11	0,23	14	0,16	0,28	0,32	0,24
3	0,34	0,08	0,26	0,32	15	0,23	0,40	0,17	0,20
4	0,42	0,25	0,15	0,18	16	0,35	0,09	0,31	0,25
5	0,29	0,16	0,31	0,24	17	0,42	0,36	0,08	0,14

6	0,13	0,09	0,41	0,37	18	0,10	0,17	0,33	0,40
7	0,33	0,24	0,27	0,16	19	0,22	0,34	0,18	0,26
8	0,40	0,31	0,19	0,10	20	0,31	0,08	0,42	0,19
9	0,08	0,17	0,42	0,33	21	0,25	0,37	0,15	0,23
10	0,25	0,42	0,18	0,15	22	0,19	0,41	0,09	0,31
11	0,19	0,26	0,31	0,24	23	0,40	0,12	0,30	0,18
12	0,30	0,08	0,22	0,40	24	0,08	0,29	0,41	0,22

5 Пример выполнения работы

Пусть при передаче сообщений по каналу связи с шумами была получена следующая статистика: символ s_1 из 100 раз был принят 97 раз, 2 раза был принят символ s_2 и 1 раз - символ s_3 ; при передаче s_2 98 раз был принят s_2 , 2 раза - s_1 , при передаче s_3 96 раз был принят s_3 , 2 раза - s_2 и 2 раза - s_4 ; при передаче s_4 99 раз был принят s_4 и 1 раз - s_3 . Вероятности появления этих символов в передаваемых сообщениях соответственно равны:

$$p(s_1) = 0,37; \quad p(s_2) = 0,3; \quad p(s_3) = 0,23; \quad p(s_4) = 0,1.$$

Канальная матрица будет иметь следующий вид:

$$P(a|b) = \begin{bmatrix} 0,97 & 0,02 & 0,01 & 0 \\ 0,02 & 0,98 & 0 & 0 \\ 0 & 0,02 & 0,96 & 0,02 \\ 0 & 0 & 0,01 & 0,99 \end{bmatrix}$$

Определим общую условную энтропию сообщений:

$$H(B|A) = - \sum_{i=1}^m \sum_{j=1}^m p(a_i) p(b_j|a_i) \log_2 p(b_j|a_i);$$

$$\begin{aligned} H(B|A) = & - [0,37 \cdot (0,97 \cdot \log_2 0,97 + 0,02 \cdot \log_2 0,02 + 0,01 \cdot \log_2 0,01) + \\ & + 0,3(0,02 \cdot \log_2 0,02 + 0,98 \cdot \log_2 0,98) + 0,23(0,02 \cdot \log_2 0,02 + 0,96 \cdot \\ & \log_2 0,96) + 0,01 \cdot (0,02 \cdot \log_2 0,02 + 0,96 \cdot \log_2 0,96) + 0,1(0,01 \cdot \log_2 0,01 + \\ & 0,99 \cdot \log_2 0,99) = 0,37 (0,043 + 0,113 + 0,066) + 0,3 (0,113 + 0,029) + \\ & 0,23 (0,113 + 0,057) + 0,01 (0,066 + 0,014) = 0,082 + 0,042 + 0,064 + \\ & 0,008 = 0,196 \text{ бит/символ.} \end{aligned}$$

Определим энтропию принятых сообщений:

$$H(B|A) = - \sum_{i=1}^m p(b_i) \log_2 p(b_i);$$

где $p(b_j) = \sum_{i=1}^m p(a_i) p(b_j|a_i)$; причём $\sum_{j=1}^m p(b_j) = 1$.

$$p(b_1) = 0,37 \cdot 0,97 + 0,3 \cdot 0,02 = 0,3649;$$

$$p(b_2) = 0,37 \cdot 0,02 + 0,3 \cdot 0,98 + 0,23 \cdot 0,02 = 0,306;$$

$$p(b_3) = 0,37 \cdot 0,01 + 0,23 \cdot 0,96 + 0,1 \cdot 0,01 = 0,2255;$$

$$p(b_4) = 0,23 \cdot 0,02 + 0,1 \cdot 0,99 = 0,1036;$$

$$0,3649 + 0,306 + 0,2255 + = 1;$$

$$H(B) = - (0,3649 \cdot \log_2 0,3649 + 0,306 \cdot \log_2 0,306 + 0,2255 \cdot \log_2 0,2255 + \\ + 0,1036 \cdot \log_2 0,1036 = 0,5306 + 0,5227 + 0,4842 + 0,3377 = 1,8352 \\ \text{бит/символ.}$$

Составим матрицу совместных вероятностей, определив для этого все совместные вероятности:

$$p(a_i, b_j) = p(a_i) \cdot p(b_j|a_i);$$

$$p(a_1 b_1) = 0,37 \cdot 0,97 = 0,3589;$$

$$p(a_2 b_3) = 0,37 \cdot 0,01 = 0,0037;$$

$$p(a_2 b_1) = 0,3 \cdot 0,02 = 0,006;$$

$$p(a_2 b_3) = 0,3 \cdot 0 = 0;$$

$$p(a_3 b_1) = 0,23 \cdot 0 = 0;$$

$$p(a_3 b_3) = 0,3 \cdot 0,96 = 0,2208;$$

$$p(a_4 b_1) = 0,1 \cdot 0 = 0;$$

$$p(a_4 b_3) = 0,1 \cdot 0,01 = 0,001;$$

$$p(a_1 b_2) = 0,37 \cdot 0,02 = 0,0074;$$

$$p(a_1 b_4) = 0,37 \cdot 0 = 0;$$

$$p(a_2 b_2) = 0,3 \cdot 0,98 = 0,249;$$

$$p(a_2 b_4) = 0,3 \cdot 0 = 0;$$

$$p(a_3 b_2) = 0,23 \cdot 0,02 = 0,0046;$$

$$p(a_3 b_4) = 0,23 \cdot 0,02 = 0,0046;$$

$$p(a_4 b_2) = 0,1 \cdot 0 = 0;$$

$$p(a_4 b_4) = 0,1 \cdot 0,99 = 0,099;$$

$$P(a,b) = \begin{bmatrix} 0.3589 & 0.0074 & 0.0037 & 0 \\ 0.006 & 0.249 & 0 & 0 \\ 0 & 0.0046 & 0.2208 & 0.0046 \\ 0 & 0 & 0.001 & 0.099 \end{bmatrix}$$

Определим энтропию объединения:

$$H(A, B) = - \sum_{i=1}^m \sum_{j=1}^m p(a_i, b_j) \log_2 p(a_i, b_j);$$

$$H(A, B) = - (0,3589 \cdot \log_2 0,3589 + 0,0074 \cdot \log_2 0,0074 + \\ + 0,00374 \cdot \log_2 0,0037 + 0,006 \cdot \log_2 0,006 + 0,2494 \cdot \log_2 0,249 + \\ + 0,0046 \cdot \log_2 0,0046 + 0,22084 \cdot \log_2 0,2208 + 0,00464 \cdot \log_2 0,0046 + \\ + 0,001 \cdot \log_2 0,001 + 0,099 \cdot \log_2 0,099) = 2,0494 \text{ бит/символ.}$$

Найдем информационные потери при передаче 500 символов:

$$\Delta I = n \cdot H(A|B);$$

$$H(A|B) = H(A, B) - H(B).$$

$$H(A|B) = 2,0494 - 1,8352 = 0,2142 \text{ бит/символ.}$$

$$\Delta I = 500 \cdot 0,2142 = 107,1 \text{ бит.}$$

Отсюда количество принятой информации будет:

$$\Delta I = n \cdot H(B) - \Delta I;$$

$$\Delta I = 500 \cdot 1,8352 - 107,1 = 810,5 \text{ бит.}$$

5 Контрольные вопросы

- 1) Что характеризует и как определяется условная энтропия одного дискретного источника сообщений относительно другого?
- 2) В каком случае условная энтропия равна нулю?
- 3) Как строится и что характеризует канальная матрица?
- 4) Что характеризует и как определяется энтропия объединения двух дискретных источников сообщений?
- 5) Чему равна энтропия объединения источников сообщений при их полной статистической зависимости?
- 6) Как связана энтропия объединения и условная энтропия двух статистически зависимых источников сообщений?
- 7) Как определяются информационные потери при передаче сообщений по каналу связи с помехами?

Практическая работа №5 «Изучение методов Шеннона – Фано и Хаффмана по построению эффективных кодов»

1 Цель работы

- изучить возможности эффективного кодирования информации по методам Шеннона – Фано и Хаффмана.

2 Основные теоретические сведения

Кодирование, при котором обеспечивается минимальная средняя длина кодовых слов, называется **эффективным** (оптимальным). В эффективном коде символу, встречающемуся чаще всего, присваивается наиболее короткая кодовая комбинация.

Задачи эффективного кодирования заключаются в следующем:

- 1) Запоминание максимального количества информации в ограниченной памяти.
- 2) Обеспечение максимальной пропускной способности канала связи.

Эффективное кодирование базируется на теореме Шеннона о кодировании при отсутствии помех, согласно которой минимальная средняя длина кодовых слов определяется соотношением:

$$\bar{L}_{\min} = \frac{H}{\log_2 K};$$

где H – энтропия источника сообщений, K – основание кода. Для двоичного кода, очевидно, что $L_{\min} = H$.

Эффективностью кода x называется отношение L_{\min} к реально достигнутой в данном коде средней длине кодовых слов \bar{L} :

$$\chi = \frac{\bar{L}_{\min}}{\bar{L}} = \frac{H}{\bar{L} \log_2 K}.$$

Средняя длина кодовых комбинаций может быть найдена следующим образом:

$$\bar{L} = \sum_{i=1}^m p(s_i) L_i;$$

где L_i - длина кодовой комбинации, соответствующей символу s_i из алфавита размером m .

Для случая отсутствия статистической взаимосвязи между символами метод построения эффективных кодов впервые был предложен Шенноном и Фано.

Для двоичного кода метод Шеннона-Фано сводится к следующему:

1) Буквы алфавита располагаются в порядке убывания вероятностей.

2) Алфавит букв разбивается на две группы таким образом, чтобы суммарные вероятности букв обеих групп были по возможности равны. Первой группе присваивается символ 1, второй символ – 0.

3) Каждую из образованных групп вновь делят на две части с приблизительно равными суммарными вероятностями и присваивают им 1 и 0. Таким образом, получают вторые цифры кода.

4) Процесс повторяется до тех пор, пока в каждой подгруппе не останется по одной букве.

Рассмотрим пример построения кода Шеннона-Фано для алфавита из шести символов (таблица 1). Кодовое дерево, соответствующее полученному коду Шеннона-Фано, представлено на рисунке 1.

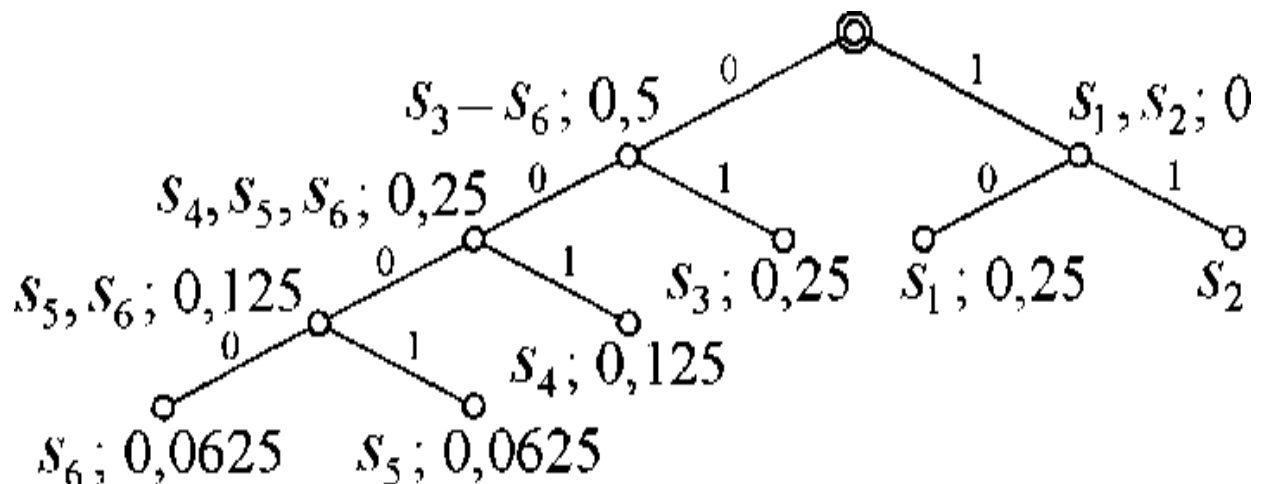


Рисунок 1 – Кодовое дерево для кода Шеннона-Фано

Таблица 1 – Получение эффективного кода по методу Шеннона – Фано

Символы исходного алфавита	P_i, s_i	Разделение символов на группы				Кодовые слова	
		1	2	3	4		
S_1	0,25	} I-1	I-1			11	
S_2	0,25		II-0			10	
S_3	0,25	} II-0	III-1			01	
S_4	0,125		} IV-0		IV-1		001
S_5	0,0625				- 0	V-1	0001
S_6	0,0625					VI-0	0000

Метод Шеннона-Фано не всегда приводит к однозначному построению кода. От указанного недостатка свободен метод Хаффмана.

Для двоичного кода метод Хаффмана сводится к следующему:

1) Буквы алфавита выписываются в столбец в порядке убывания вероятностей.

2) Две последние буквы объединяются в одну вспомогательную букву, которой приписывается суммарная вероятность.

3) Вероятности букв, участвующих в объединении и полученная суммарная вероятность вновь располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние буквы объединяются.

4) Процесс продолжается до тех пор, пока не будет получена единственная вспомогательная буква с суммарной вероятностью, равной 1.

Для получения кодовой комбинации, соответствующей данной букве необходимо проследить путь перехода по строкам и столбцам таблицы.

Рассмотрим пример построения кода Хаффмана для алфавита из восьми символов (таблица 2). Результат показан на рисунке 2 в виде кодового дерева соответствующего коду Хаффмана.

Таблица 2 – Получение эффективного кода по методу Хаффмана

Символы	Вероятности	Вспомогательные столбцы			
		1	2	3	4
s_1	0,4	0,4	0,4	0,6	1
s_2	0,25	0,25	0,35		
s_3	0,2	0,2	0,25	0,4	
s_4	0,1	0,15			
s_5	0,05				

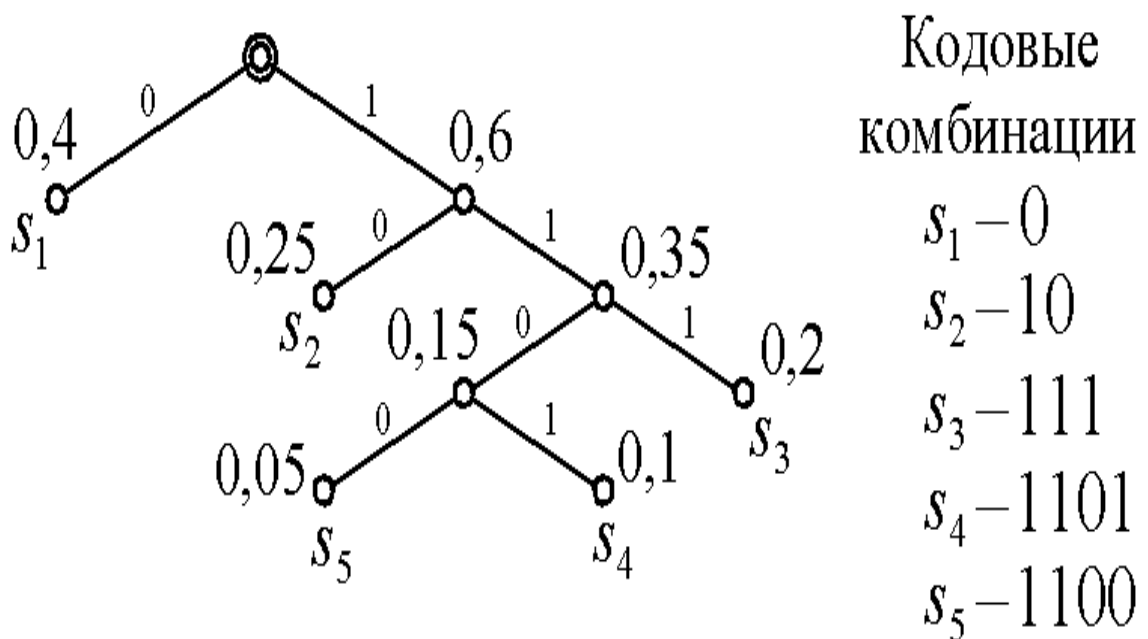


Рисунок 2 – Пример кодового дерева для кода Хаффмана

3 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.

3) На основе заданного первичного алфавита и вероятностей появления символов этого алфавита (табл. 3.3) получить в форме таблицы двоичный код Шеннона-Фано.

4) Построить кодовое дерево для полученного кода Шеннона-Фано.

5) Определить эффективность кода, полученного по методу Шеннона-Фано.

6) Выполнить пункты 3, 4, 5 для метода Хаффмана.

7) Сравнить эффективности методов Шеннона-Фано и Хаффмана. Сделать выводы о полученных результатах.

8) Оформить и защитить отчет по выполнению практической работе.

4 Варианты заданий

Номер варианта студента определяется как порядковый номер в журнале преподавателя.

Таблица 3 – Вероятности появления символов для различных вариантов

Вар.	Символы алфавита источника сообщений											
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	0,14	0,06	0,05	0,08	0,13	0,04	0,01	0,09	0,15	0,02	0,11	0,12
2	0,11	0,05	0,09	0,10	0,12	0,03	0,02	0,08	0,15	0,07	0,14	0,04
3	0,13	0,07	0,05	0,06	0,15	0,04	0,11	0,02	0,12	0,16	0,08	0,01
4	0,02	0,11	0,12	0,01	0,09	0,15	0,08	0,13	0,04	0,14	0,06	0,05
5	0,07	0,14	0,04	0,02	0,08	0,15	0,10	0,12	0,03	0,11	0,05	0,09
6	0,16	0,08	0,01	0,11	0,02	0,12	0,06	0,15	0,04	0,13	0,07	0,05
7	0,01	0,09	0,15	0,02	0,11	0,12	0,14	0,06	0,05	0,08	0,13	0,04
8	0,02	0,08	0,15	0,07	0,14	0,04	0,13	0,07	0,05	0,06	0,15	0,04
9	0,11	0,02	0,12	0,16	0,08	0,01	0,07	0,05	0,13	0,06	0,15	0,04
10	0,06	0,05	0,14	0,13	0,04	0,08	0,15	0,01	0,09	0,12	0,02	0,11
11	0,09	0,11	0,05	0,03	0,10	0,12	0,15	0,02	0,08	0,14	0,04	0,07
12	0,05	0,13	0,07	0,15	0,04	0,06	0,02	0,12	0,11	0,04	0,07	0,14
13	0,12	0,02	0,11	0,09	0,15	0,01	0,13	0,04	0,08	0,06	0,05	0,14
14	0,14	0,04	0,07	0,15	0,02	0,08	0,03	0,10	0,12	0,09	0,11	0,05
15	0,04	0,07	0,14	0,02	0,12	0,11	0,15	0,04	0,06	0,05	0,13	0,07
16	0,05	0,07	0,13	0,04	0,15	0,06	0,12	0,02	0,11	0,01	0,08	0,16
17	0,1	0,03	0,05	0,09	0,14	0,04	0,01	0,08	0,16	0,04	0,12	0,14
18	0,12	0,07	0,08	0,11	0,16	0,01	0,04	0,06	0,13	0,09	0,1	0,03
19	0,11	0,08	0,07	0,04	0,14	0,05	0,13	0,02	0,1	0,15	0,09	0,02
20	0,03	0,12	0,14	0,02	0,08	0,15	0,1	0,11	0,03	0,12	0,05	0,05
21	0,08	0,13	0,05	0,01	0,06	0,14	0,11	0,13	0,06	0,1	0,04	0,09
22	0,15	0,09	0,02	0,13	0,02	0,1	0,08	0,16	0,01	0,12	0,06	0,06
23	0,03	0,11	0,16	0,05	0,14	0,15	0,09	0,01	0,04	0,07	0,13	0,02
24	0,06	0,01	0,12	0,09	0,16	0,02	0,11	0,03	0,08	0,05	0,15	0,12

5 Контрольные вопросы

1. В чем заключается сущность эффективного кодирования?
2. Каковы основные задачи эффективного кодирования?
3. Как определяется средняя длина кодового слова?
4. Чему равна нижняя граница эффективного кодирования?
5. Как определяется эффективность кода?
6. В каком случае метод Шеннона-Фано гарантированно обеспечивает получение эффективного кода?
7. С помощью какой операции в методе Хаффмана обеспечивается получение вспомогательных символов?

Практическая работа №6
«Изучение метода арифметического кодирования
последовательностей символов»

1 Цель работы

- приобрести умение производить арифметическое кодирование последовательностей символов с целью сжатия информации, а также умение декодировать арифметический код.

2 Основные теоретические сведения

Арифметическое кодирование представляет собой метод сжатия информации, в котором кодируемая последовательность символов представляется в виде дробного числа x , принимающего значения из интервала $0 < x < 1$.

Пусть задан дискретный источник сообщений $[S, p(si)]$ с алфавитом $S = \{s_1, \dots, s_m\}$ и распределением вероятностей $p(si)$. На выходе источника задана последовательность символов $\bar{s} = (s_{i_1}, s_{i_2}, \dots, s_{i_k}, \dots, s_{i_M})$;

где i - порядковый номер символа в алфавите; k - порядковый номер символа в последовательности.

Для заданной последовательности s требуется получить кодовое слово: $\bar{x} = (x_{j_1}, x_{j_2}, \dots, x_{j_L})$; где $x_j = 0$ или 1 .

Определим величины, называемые кумулятивными вероятностями символов S :

$$q(s_1) = 0; \quad q(s_2) = p(s_1); \quad \dots \quad q(s_m) = \sum_{i=1}^{m-1} p_i.$$

В рекурсивной форме данные вероятности можно записать следующим образом:

$$q(s_i) = q(s_{i-1}) + p(s_{i-1}).$$

Обозначим через \bar{s}_{jk} последовательность символов, полученную путем присоединения к исходной последовательности $\bar{s}_{j,k-1}$, символа s_i .

Определим величины $F(\bar{s}_{jk})$ и $G(s_u)$, представляющие собой нижнюю границу и длину интервала вероятностей, который соответствует последовательности \bar{s}_{jk} :

$$F(\bar{s}_{jk}) = F(\bar{s}_{jk-1}) + q(s_i) \cdot G(\bar{s}_{jk-1})$$

$$G(\bar{s}_{jk}) = p(s_i) \cdot G(\bar{s}_{jk-1}).$$

При $k=0$ (начальный шаг алгоритма) принимают $F=0$, $G=1$. Длина кодового слова x , обеспечивающая однозначное декодирование последовательности символов определяется по формуле:

$$L = \lceil -\log_2 G(\bar{s}) \rceil + 1;$$

где \bar{s} – округление числа, стоящего в скобках, до целого в большую сторону.

Искомое кодовое слово \bar{x} определяется как L знаков после запятой в двоичной записи числа $F(\bar{s}) + G(\bar{s})/2$, где $F(s)$ и $G(s)$ – границы интервала вероятностей, соответствующие заданной последовательности \bar{s} :

$$\text{Bin}[F(\bar{s}) + G(\bar{s})/2] = 0, x_{j_2}, \dots, x_{j_L} \rightarrow \bar{x} = (x_{j_1}, x_{j_2}, \dots, x_{j_L})$$

Таким образом, метод арифметического кодирования заключается в последовательном нахождении величин $q(s_i)$, $F(s_n)$, $G(S_{jk})$, для $i = 2, 3, \dots, m$ и $k = 1, 2, \dots, M$, пока не будут определены значения $F(s)$ и $G(s)$. На основе последних значений $F(s)$ и $G(s)$ и определяется искомое кодовое слово.

Для пояснения метода арифметического кодирования можно прибегнуть к его графической интерпретации. В этом случае интервал $[0, 1)$ вероятностей появления символов рассматривается в виде отрезка длиной 1. При этом символам на данном отрезке будут соответствовать непересекающиеся полуинтервалы длинами, равными вероятностям появления символов (рисунок 1). Кумулятивные вероятности соответствуют началам этих полуинтервалов. Эти точки идентифицируют сообщения источника.

На каждом шаге алгоритма производится пересчет границ всего отрезка, в котором будет расположено число, соответствующее кодовому слову. Число $F(S_j)$ в этом случае является начальной точкой отрезка на шаге k , а число $G(\bar{s}_{ik})$ является длиной данного отрезка.

Декодирование полученного арифметического кода заключается в выполнении следующих действий. На нулевом шаге ($k=0$) принимаем $q(s_{M+1}) = 1$, $F(\bar{s}_{ik}) = 0$, $G(\bar{s}_{ik}) = 1$. Последовательно

(от $i = 1$ до M) для символов s_i на каждом шаге производим проверку выполнения условия:

$$F(\bar{s}_{ik}) + q(s_i) - G(\bar{s}_{ik}) < x.$$

Если для символа s_i данное условие не выполняется, то в качестве символа декодируемой последовательности на данном шаге принимаем s_{i-1} . В обратном случае (условие выполняется для всех i) принимаем символ s_m .

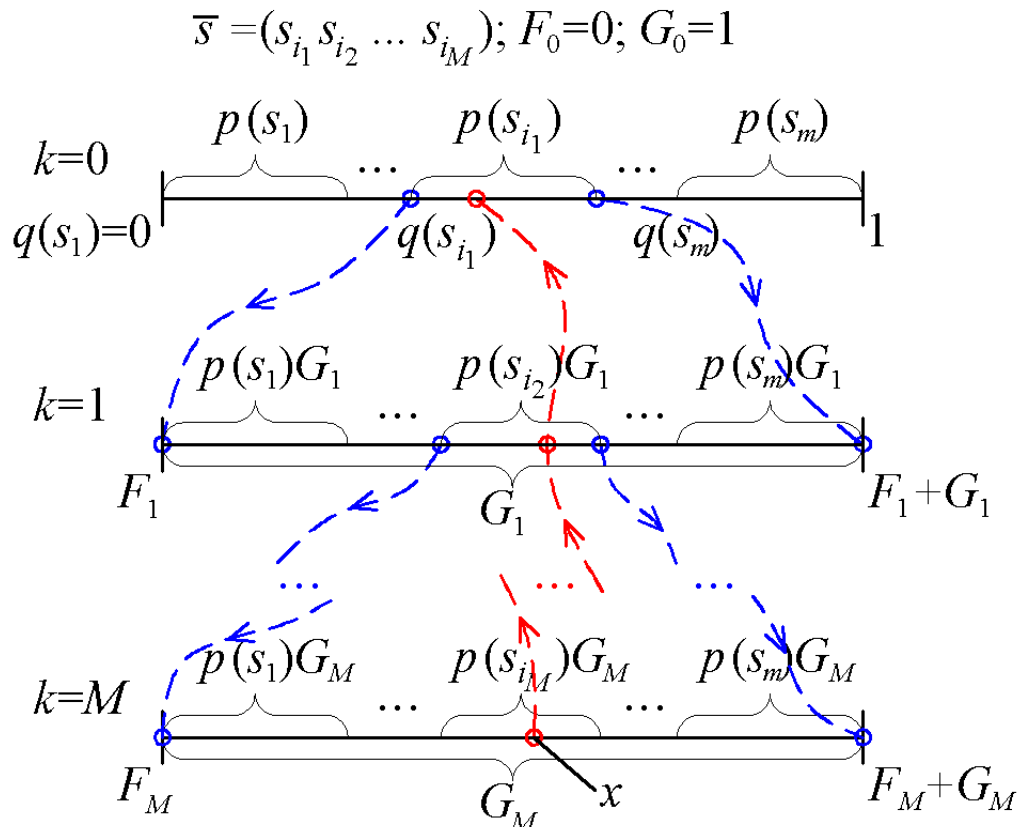


Рисунок 1 - Графическая интерпретация арифметического кодирования

Для найденного символа определяют величины $F(\bar{s}_{ik})$ и $G(\bar{s}_{ik})$ по тем же формулам, что и при кодировании.

Алгоритм продолжают до тех пор, пока не будет найден последний символ ($k = M$) декодируемой последовательности \bar{s} .

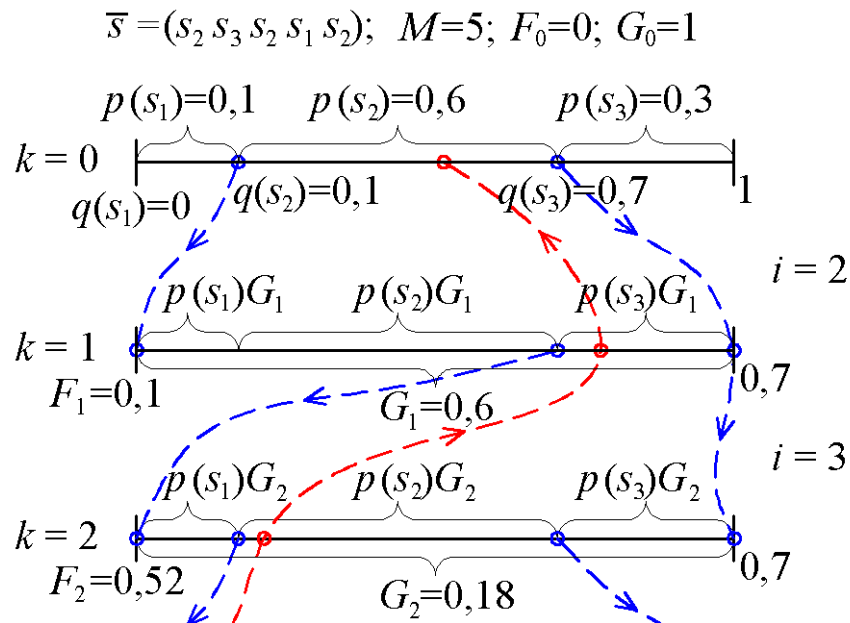
3 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) Произвести арифметическое кодирование заданной последовательности символов (таблица 1)
- 4) Построить графическую интерпретацию арифметического кодирования.
- 5) Осуществить декодирование полученного арифметического кода.
- 6) Оформить и защитить отчет по практической работе.

4 Пример выполнения работы

Пусть задан дискретный источник сообщений с алфавитом $S=\{s_1, s_2, s_3\}$ и распределением вероятностей $p(s_1)=0,1$, $p(s_2)=0,6$, $p(s_3)=0,3$. Произведем арифметическое кодирование последовательности $\bar{s} = (s_2 s_3 s_2 s_1 s_2)$ длиной $M=5$. Процесс арифметического кодирования последовательности s представлен в таблице 1. Графическая интерпретация арифметического кодирования показана на рисунке 2.



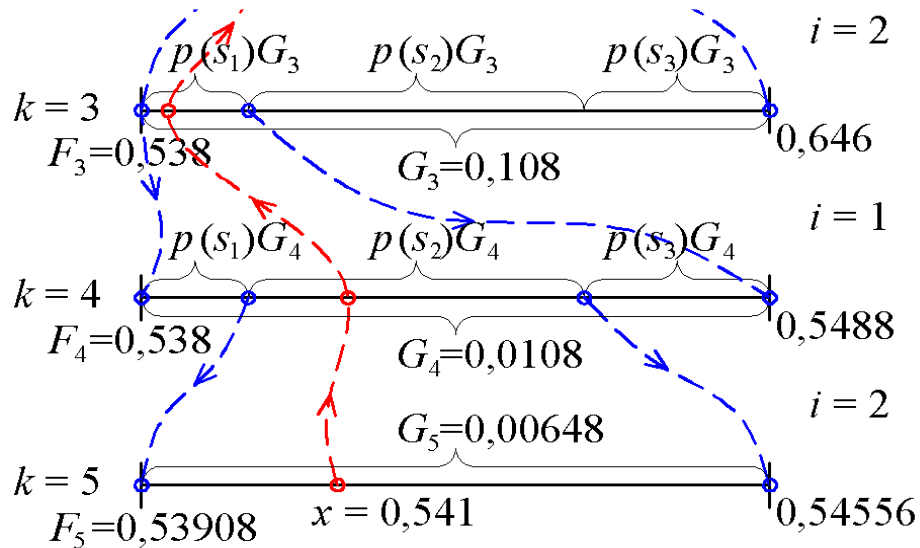


Рисунок 2 – Графическая интерпретация арифметического кодирования последовательности $\bar{s} = (s_2s_3s_2s_1s_2)$

Длина кодового слова будет $L = \lceil -\log_2 0.00648 \rceil + 1 = \lceil 7.2698 \rceil + 1 = 9$, кодовом слове должно быть не меньше 9 знаков.

Таблица 1 – Арифметическое кодирование последовательности

Шаг k ;	S_i	\bar{s}_{ik}	$p(s_i)$	$q(s_i)$	$F(\bar{s}_{ik})$	$G(\bar{s}_{ik})$
0	—	—	—	—	0	1
1	S_2	S_2	0,6	0,1	0,1	0,6
2	S_3	S_2S_3	0,3	0,7	0,52	0,18
3	S_2	$S_2S_3S_2$	0,6	0,1	0,538	0,108
4	S_1	$S_2S_3S_2S_1$	0,1	0	0,538	0,0108
5	S_2	$S_2S_3S_2S_1S_2$	0,6	0,1	0,53908	0,00648

Отсюда кодовое слово запишется следующим образом $x = \text{bin}(0,53908 + 0,00648/2) = \text{bin}(0,54232) = 0,100010101_2$. Процесс декодирования арифметического кода $x = 100010101$ представлен в таблице 2.

Таблица 2 – Декодирование арифметического кода $x = 100010101$

Шаг к	F_k	G_k	Гипоте за S_i	$q(s_i)$	Проверка $F_k + q_i G_k < x$	Решение S_i	$p(s_i)$
0	$x = 100010101 \rightarrow + x = 0.100010101_2 = 0,541_{10}$						
1	0	1	S_1	0	$0 < x$	S_2	0,6
			S_2	0Д	$0Д < x$		
			S_3	0,7	$0,7 > x$		
2	0,1	0,6	S_1	0	$0Д < x$	S_3	0,3
			S_2	од	$0,16 < x$		
			S_3	0,7	$0,52 < x$		
3	0,52	0,18	S_1	0	$0,52 < x$	S_2	0,6
			S_2	од	$0,538 < x$		
			S_3	0,7	$0,646 > x$		
4	0,538	0,108	S_1	0	$0,538 < x$	S_1	ОД
			S_2	од	$0,5488 > x$		
5	0,538	0,0108	S_i	0	$0,538 < x$	S_2	0,6
			S_3	0,7	$0,5450 > x$		

Отсюда последовательность символов, полученная путем декодирования арифметического кода, будет $S_1 S_3 S_2 S_1 S_2$.

5 Варианты заданий

Требуется произвести арифметическое кодирование последовательности символов длиной $M=7$ и декодирование полученного кодового слова. Кодированные последовательности символов и вероятности их появления для каждого варианта задания представлены в таблице 3.

Номер варианта определяется как порядковый номер студента в журнале преподавателя.

Таблица 3 – Кодированная последовательность символов и вероятности их появления

Var.	S	$p(S_i)$	Var.	s	$p(s_i)$
1	$S_2 S_4 S_1 S_2 S_3 S_1 S_4$	$p(s_1) = 0,1$	13	$S_3 S_1 S_2 S_3 S_1 S_3 S_4$	$p(s_1) = 0,3$
2	$S_4 S_2 S_1 S_3 S_2 S_4 S_1$	$p(s_2) = 0,4$	14	$S_1 S_3 S_1 S_2 S_3 S_4 S_3$	$p(s_2) = 0,2$
3	$S_1 S_4 S_3 S_2 S_4 S_1 S_2$	$p(s_3) = 0,2$	15	$S_4 S_3 S_1 S_3 S_1 S_2 S_3$	$p(s_3) = 0,4$
4	$S_4 S_1 S_2 S_3 S_1 S_2 S_4$	$p(s_4) = 0,3$	16	$S_3 S_4 S_3 S_1 S_3 S_2 S_1$	$p(s_4) = 0,1$
5	$S_1 S_3 S_1 S_4 S_3 S_2 S_1$	$P(s_i) = 0,4$	17	$S_2 S_3 S_1 S_2 S_4 S_4 S_2$	$p(s_1) = 0,2$
6	$S_3 S_1 S_4 S_1 S_2 S_1 S_3$	$p(s_2) = 0,1$	18	$S_4 S_2 S_2 S_3 S_1 S_2 S_4$	$p(s_2) = 0,4$

7	$S4S3S1S2S1S3S1$	$p(s_3) = 0,3$	19	$S4S3S2S4S2S2S1$	$p(s_3) = 0,1$
8	$S1S4S2S1S3S1S3$	$P(s_4) = 0,2$	20	$S3S2S4S4S2S1S2$	$P(s_4) = 0,1$
9	$S4S2S4S1S3S4S2$	$P(s_i) = 0,2$	21	$S2S1S3S2S3S3S4$	$p(s_1) = 0,1$
10	$S2S4S1S4S2S4S3$	$p(s_2) = 0,3$	22	$S3S2S2S3S4S3S1$	$p(s_2) = 0,3$
11	$S1S4S2S3S4S2S4$	$P_{iss} = 0,1$	23	$S1S2S3S3S2S4S3$	$p(s_3) = 0,4$
12	$S4S1S4S2S4S3S2$	$P(s_4) = 0,4$	24	$S3S1S3S4S2S2S3$	$p(s_4) = 0,2$

6 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) На основе заданного первичного алфавита и вероятностей появления символов этого алфавита (табл. 3.3) получить в форме таблицы двоичный код Шеннона-Фано.
- 4) Построить кодовое дерево для полученного кода Шеннона-Фано.
- 5) Определить эффективность кода, полученного по методу Шеннона-Фано.
- 6) Выполнить пункты 3, 4, 5 для метода Хаффмана.
- 7) Сравнить эффективности методов Шеннона-Фано и Хаффмана. Сделать выводы о полученных результатах.
- 8) Оформить и защитить отчет по выполнению практической работы.

7 Контрольные вопросы

- 1) В какую форму преобразуется последовательность символов в методе арифметического кодирования?
- 2) Что понимают под кумулятивными вероятностями символов?
- 3) Как в методе арифметического кодирования определяются значения величин F_k и G_k ?
- 4) Что представляют собой величины $p(s_i)$, $q(s_i)$, $F_k \cdot G_k$ в графической интерпретации арифметического кодирования?

- 5) Как находят значение числа, соответствующего арифметическому коду?
- 6) В чем заключается декодирование арифметического кода?
- 7) Какова графическая интерпретация декодирования арифметического кода?

Практическая работа №7 «Исследование линейных блочковых кодов»

1 Цель работы

- приобрести умение строить линейные блочковые коды на основе порождающих матриц для обнаружения и исправления ошибок в кодовых словах.

2 Основные теоретические сведения

Блочковыми называют помехоустойчивые коды, в которых процедура кодирования заключается в разбиении входной последовательности информационных символов на блоки, содержащие m символов. Каждому информационному блоку длиной m сопоставляется k проверочных символов. Полученное кодовое слово из $n = m + k$ символов называют кодовым блоком.

Число несовпадающих позиций в двух кодовых словах x и y называется *расстоянием Хэмминга* $d(x, y)$ между этими словами.

Для двоичных кодовых слов расстояние Хэмминга может быть получено как число единиц в сумме в кодовых словах по модулю 2. Правила сложения по модулю 2 определяются следующим образом:

$$0 \oplus 0 = 0; 0 \oplus 1 = 1; 1 \oplus 0 = 1; 1 \oplus 1 = 0.$$

Важной характеристикой корректирующего блочкового кода C является *кодвое расстояние*, которое принимается равным наименьшему расстоянию Хэмминга между словами данного кода:

$$d(C) = \min\{d(\bar{x}, \bar{y}) : \bar{x}, \bar{y} \in C, \bar{x} \neq \bar{y}\}.$$

Для блочковых кодов справедливы следующие утверждения:

1) Для того чтобы блочковый код C позволял обнаруживать все комбинации из t или менее ошибок, необходимо и достаточно, чтобы его кодвое расстояние было равно $d(C) = t + 1$.

2) Для того чтобы блочковый код C позволял исправлять все комбинации из t или менее ошибок необходимо и достаточно, чтобы его кодвое расстояние было равно $d(C) = 2t + 1$.

Для практических расчетов при определении числа проверочных символов k в коде с кодовым расстоянием $d = 3$ используют следующие формулы:

если известна длина полного кодового слова n , то:

$$k = \lceil \log_2(n+1) \rceil;$$

если при расчетах удобнее исходить из заданного числа информационных символов m , то:

$$k = \lceil \log_2 \{ (m+1) + \lceil \log_2(m+1) \rceil \} \rceil;$$

где $\lceil \dots \rceil$ - округление числа, стоящего в скобках, до целого в большую сторону.

Для блочных кодов с $d = 4$

$$k \geq 1 + \log_2(n+1)$$

или

$$k \geq 1 + \lceil \log_2 \{ (m+1) + \log_2(m+1) \} \rceil.$$

Самый большой класс блочных корректирующих кодов составляют **линейные коды**, у которых значения проверочных символов определяются при проведении линейных операций над информационными символами.

К линейным операциям относятся сложение и умножение на постоянное число. Для двоичных линейных кодов в качестве линейной операции используют сложение по модулю 2.

Кодовые комбинации линейных блочных кодов называют **кодowymi векторами**. Кодовый вектор, состоящий из одних нулей, называется **нулевым вектором**. Число единиц в двоичном кодовом векторе называется **весом кодового вектора**.

Кодовое расстояние для линейного блочного кода равно минимальному весу его ненулевых кодовых векторов.

Линейные блочные коды задают при помощи **порождающих матриц**, размерность которых равна $m \times n$:

$$G_{m,n} = [E_m | P_{m,k}] = \left[\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & p_{11} & p_{12} & \dots & p_{1k} \\ 1 & 0 & \dots & 0 & p_{21} & p_{22} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 & p_{m1} & p_{m2} & \dots & p_{mk} \end{array} \right]$$

где E_m - единичная матрица размера m ; $P_{m,k}$ - матрица-дополнение, состоящая из проверочных символов p_{ij} .

Для кодов с $d = 2$ порождающая матрица имеет следующий вид:

$$G_{m,n} = \left[\begin{array}{cccc|c} 1 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 & 1 \end{array} \right]$$

Во всех кодовых векторах, построенных при помощи такой матрицы, будет четное число единиц.

Для кодов с $d > 3$ вид матрицы P зависит от конкретных требований к порождаемому коду. Этими требованиями могут быть либо минимум проверочных символов, либо максимальная простота кодирующей/декодирующей аппаратуры.

Линейные блочные коды называются *совершенными* (плотно упакованными), если они обнаруживают и исправляют максимальное число ошибок при минимальном числе проверочных символов.

Чем больше вес строк матрицы P , тем ближе порождаемый код к совершенному. С другой стороны, число единиц в матрице P определяет число сумматоров по модулю 2 в кодере и декодере канала связи. То есть, чем больше единиц в матрице P , тем сложнее аппаратура.

При построении совершенных кодов с $d > 3$ последовательно используются строки матрицы P с весом $W_p = k, k-1, k-2, \dots, d-1$. Таким образом, вес W_p каждой строки матрицы P должен удовлетворять условию:

$$W_p \geq d - 1$$

При построении кодов с максимально простыми кодерами и декодерами каналов связи в матрице P последовательно выбираются строки весом $W_p = 2, 3, \dots, k$.

Если число комбинаций строк матрицы P , удовлетворяющих условию $W_p \geq d - 1$, больше t , то в первом случае (получение совершенного кода) не используют комбинации с наименьшим весом, а во втором (максимальная простота аппаратуры) - с наибольшим.

Матрицу P следует строить таким образом, чтобы число единиц в столбцах данной матрицы было по возможности одинаковым. От числа единиц в столбце матрицы P зависит число проверок, производимых при исправлении ошибок.

В общем виде формулы для определения значений

проверочных символов могут быть записаны следующим образом:

$$p_j = \sum_{i=1}^m p_{ij} x_i; \quad (j=1, 2, \dots, k);$$

где p_{ij} - элементы матрицы-дополнения P .

Для двоичных кодов значения проверочных символов p_j находят путем суммирования по модулю 2 тех строк матрицы P , номера которых совпадают с номерами информационных символов, значение которых равно единице.

Для линейного блочного кода C с порождающей матрицей $G_{m,n} = [E_m | P_{m,k}]$ проверочной матрицей будет являться матрица:

$$I'_{k,n} = [P_{m,k}^T | E_k]$$

Число проверок равно числу k проверочных символов кода.

В общем виде систему проверок можно записать следующим образом:

$$p_j \oplus \sum_{i=1}^m p_{ij} a_i = S_j; \quad j=1, 2, \dots, k.$$

В результате осуществления проверок образуется проверочный вектор $S = [S_1 S_2 \dots S_j \dots S_k]$, называемый *синдромом*. Если все разряды синдрома равны нулю, то принятый кодовый вектор считается безошибочным. Если хотя бы один из разрядов синдрома содержит единицу, то принятый кодовый вектор содержит ошибку.

Исправление одиночной ошибки осуществляется по виду синдрома, так как каждому ошибочному разряду принятого кодового вектора соответствует один единственный синдром. Вид синдрома может быть определен при помощи проверочной матрицы H . Столбцы матрицы H представляют собой синдром для соответствующего ошибочного разряда кодового вектора.

В общем случае для исправления заданного числа ошибок t с помощью линейных блочных кодов используют *стандартные таблицы декодирования*, содержащие все возможные значения принятых из канала векторов z . Данные таблицы организованы таким образом, чтобы мог быть найден ближайший к z переданный кодовый вектор y (таблица 1).

В первой строке таблицы располагаются все кодовые векторы y_i , число которых составляет $N = 2^m$. В первом столбце второй строки размещается вектор ошибки \bar{e}_1 , вес которого равен 1.

Остальные ячейки второй строки заполняются векторами, полученными в результате суммирования по модулю 2 вектора \bar{e}_1 с вектором \bar{y}_i расположенным в соответствующем столбце первой строки. В первом столбце третьей строки записывается вектор \bar{e}_2 , вес которого также равен 1, однако, если вектор \bar{e}_1 содержит единицу в первом разряде, то \bar{e}_2 - во втором. В остальные ячейки третьей строки записывают суммы \bar{x}_i , и \bar{e}_2 .

Таблица 1 – Стандартная таблица для декодирования линейного блочного кода

\bar{s}_j	\bar{e}_j	\bar{y}_1	\bar{y}_2	...	\bar{y}_N
\bar{s}_1	\bar{e}_1	$\bar{y}_1 + \bar{e}_1$	$\bar{y}_2 + \bar{e}_1$...	$\bar{y}_N + \bar{e}_1$
\bar{s}_2	\bar{e}_2	$\bar{y}_1 + \bar{e}_2$	$\bar{y}_2 + \bar{e}_2$...	$\bar{y}_N + \bar{e}_2$
...			
\bar{s}_K	\bar{e}_K	$\bar{y}_1 + \bar{e}_K$	$\bar{y}_2 + \bar{e}_K$...	$\bar{y}_N + \bar{e}_K$

Аналогично поступают до тех пор, пока не будут просуммированы с векторами \bar{y}_i все векторы \bar{e}_j весом 1, с единицами в каждом из n разрядов. Затем суммируются по модулю 2 векторы \bar{e}_j весом 2, с последовательным перекрытием всех возможных разрядов. Вес вектора \bar{e}_j определяет число исправляемых ошибок. Число векторов \bar{e}_j определяется возможным числом неповторяющихся синдромов и равно $K = 2^k - 1$ (нулевая комбинация говорит об отсутствии ошибки). Условие неповторяемости синдрома позволяет по его виду определять соответствующий ему вектор \bar{e}_j .

По виду синдрома \bar{s}_j принятый кодовый вектор может быть отнесен к определенной строке таблицы соответствия (смежному классу). Принятый кодовый вектор сравнивается с векторами, записанными в данную строку и в случае совпадения в каком-либо из столбцов выбирается вектор (истинный), расположенный в первой строке данного столбца.

Векторы ошибок \bar{e}_j не должны быть равны ни одному из кодовых векторов.

3 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) Построить производящую матрицу G для линейного блочного кода, способного исправлять одиночную ошибку.
- 4) Для заданных последовательностей информационных символов получить кодовые вектора линейного блочного кода.
- 5) На основе матрицы-дополнения P к порождающей матрице G получить систему проверок для нахождения синдрома.
- 6) Построить проверочную матрицу линейного блочного кода.
- 7) Показать процесс исправления одиночной ошибки в произвольном разряде полученных кодовых векторов на основе синдрома и проверочной матрицы.
- 8) Построить стандартную таблицу декодирования линейного блочного кода, позволяющую исправить максимально возможное число двойных ошибок.
- 9) Показать процесс исправления двойной ошибки в произвольном разряде одного из кодовых векторов.
- 10) Оформить и защитить отчет по выполнению практической работы.

4 Пример выполнения работы

Пусть требуется построить порождающую матрицу для линейного блочного кода, способного исправлять одиночную ошибку при передаче информационных векторов из 7 символов (например, 0100111, 1101100, 1011010).

Получим порождающую матрицу в приведенной форме:

$$G_{m,n} = [E_m | P_{m,k}].$$

Поскольку число исправляемых ошибок $t = 1$, то кодовое расстояние для линейного блочного кода будет $d = 2t + 1$; $d = 2 * 1 + 1 = 3$.

Так как длина информационных векторов $m = 7$, то число

строк порождающей матрицы линейного блочного кода должно быть равно 7. Число столбцов порождающей матрицы равно длине кодовых векторов

$$n = m + k;$$

где k - число проверочных символов, которое при $d = 3$ может быть найдено по формуле:

$$k = \left\lceil \log_2 \left\{ (m+1) + \left\lceil \log_2 (m+1) \right\rceil \right\} \right\rceil$$

$$k = \left\lceil \log_2 \left\{ (7+1) + \left\lceil \log_2 (7+1) \right\rceil \right\} \right\rceil = \left\lceil \log_2 (8+3) \right\rceil = 4$$

Отсюда получим $n = 7 + 4 = 11$.

Таким образом, искомым линейный блочный код является $(11,7,3)$ – кодом.

Поскольку вес каждой строки матрицы-дополнения P должен быть $W_P > d - 1$, то в качестве строк матрицы P примем четырехзначные двоичные комбинации с числом единиц $W_P > 2$. Комбинации выберем так, чтобы число единиц в столбцах матрицы P было одинаковым. Заданным требованиям удовлетворяет следующий набор: 1111, 1110, 1101, 1011, 0111, 0110, 1001. Окончательный вид порождающей матрицы будет:

$$G_{7,11} = \left[\begin{array}{cccccc|ccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{array} \right]$$

Получим для заданных информационных векторов 0100111, 1101100, 1011010 кодовые вектора линейного $(11,7,3)$ -кода. Для этого найдем значения проверочных символов для каждого информационного вектора путем суммирования по модулю 2 тех строк матрицы P , номера которых совпадают с номерами разрядов, содержащих единицы в информационных векторах:

$$\begin{array}{rcl}
 1) \underline{0100111}: & 2) \underline{1101100}: & 3) \underline{1011010}: \\
 1110 & 1111 & 1111 \\
 \oplus 0111 & \oplus 1110 & \oplus 1101 \\
 0110 & 1011 & 1011 \\
 \underline{1001} & \underline{0111} & \underline{0110} \\
 0110 & 1101 & 1111
 \end{array}$$

Отсюда искомые кодовые векторы будут: $\bar{x}_1 = 0100110110$; $\bar{x}_2 = 11011001101$; $\bar{x}_3 = 10110101111$.

Запишем систему проверок для найденной матрицы P :

$$p_j \oplus \sum_{i=1}^m p_{ij} x_i = S_j; (j = 1, 2, \dots, k).$$

Система проверок будет иметь следующий вид:

$$\begin{cases}
 p_1 \oplus x_1 \oplus x_2 \oplus x_3 \oplus x_4 \oplus x_7 = S_1, \\
 p_2 \oplus x_1 \oplus x_2 \oplus x_3 \oplus x_5 \oplus x_6 = S_2, \\
 p_3 \oplus x_1 \oplus x_2 \oplus x_4 \oplus x_5 \oplus x_6 = S_3, \\
 p_4 \oplus x_1 \oplus x_3 \oplus x_4 \oplus x_5 \oplus x_7 = S_4.
 \end{cases}$$

Для того чтобы знать, какая комбинация синдрома S будет соответствовать ошибке в определенном разряде принятого кодового вектора, построим проверочную матрицу линейного блокового кода:

$$I'_{k,n} = [P_{m,k}^T | E_k]$$

Транспонированная матрица-дополнение P будет выглядеть следующим образом:

$$P_{7;4}^T = \begin{bmatrix} 1111011 \\ 1110110 \\ 1101100 \\ 1011101 \end{bmatrix}$$

Отсюда проверочная матрица линейного блокового кода будет

$$H_{4;11} = \begin{array}{c|cccccccc|cccc}
 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & P_1 & P_2 & P_3 & P_4 \\
 \hline
 1 & 1 & 1 & 1 & 0 & 0 & 1 & & 1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 1 & 1 & 0 & & 0 & 1 & 0 & 0 \\
 1 & 1 & 0 & 1 & 1 & 1 & 0 & & 0 & 0 & 1 & 0
 \end{array}$$

$$\left| \begin{array}{cccccc|cccc} 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array} \right|$$

Предположим, что в переданных кодовых векторах изошли следующие ошибки (выделены полужирным шрифтом):
 $x_1 = 01001010110$; $x_2 = 11111001101$; $x_3 = 10110101101$.

Найдем согласно полученной системе проверок для каждого переданного кодового вектора синдром $S = [S_1 S_2 \dots S_k]$:

1). Для кодового вектора X_1 :

X_1	X_2	X_3	X_4	X_5	X_6	X_7	P_1	P_2	P_3	P_4
0	1	0	0	1	0	1	0	1	1	0

$$\begin{cases} 0 \oplus 0 \oplus 1 \oplus 0 \oplus 0 \oplus 1 = 0, \\ 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 = 1, \\ 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 = 1, \\ 0 \oplus 0 \oplus 0 \oplus 0 \oplus 1 \oplus 1 = 0. \end{cases}$$

Синдром 0110 показывает, что значение символа x_6 следует заменить на противоположное.

2). Для кодового вектора X_2 :

X_1	X_2	X_3	X_4	X_5	X_6	X_7	P_1	P_2	P_3	P_4
1	1	1	1	1	0	0	1	1	0	1

$$\begin{cases} 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 = 1, \\ 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 = 1, \\ 0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 = 0, \\ 1 \oplus 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 = 1. \end{cases}$$

Синдром 0110 показывает, что ошибка произошла в символе X_3 .

1) Для кодового вектора X_3 :

X_1	X_2	X_3	X_4	X_5	X_6	X_7	P_1	P_2	P_3	P_4
1	0	1	1	0	1	0	1	1	0	1

$$\begin{cases} 1 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 0, \\ 1 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 = 0, \\ 0 \oplus 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 = 1, \\ 1 \oplus 1 \oplus 1 \oplus 1 \oplus 0 \oplus 0 = 0. \end{cases}$$

Синдром 0010 показывает, что значение символа p_3 является ошибочным.

5 Варианты заданий

Номер варианта определяется как порядковый номер студента в журнале преподавателя

Таблица 2 – Варианты заданий для построения линейных блочковых кодов

Вариант	Информационные векторы		
1	10111010	11010101	10101011
2	00010100	00001101	00100010
3	10010100	00110001	10011101
4	00011110	01100000	00100010
5	11100011	00011111	10011101
6	01001010	00110010	01000100
7	10001011	00001101	11010001
8	00010101	10001110	01111001
9	11011110	10110001	10100001
10	01000010	10011110	01011101
11	11000011	11000001	10011101
12	00011010	11010100	00101010
13	10000111	11000110	10010001
14	00001110	10011100	00010011
15	10011100	00110101	11011101
16	00101100	00101000	11000100
17	00010111	11101110	00010001
18	10110100	01011001	11000100
19	00010011	10000110	11100011
20	00101110	00100100	01101000
21	10101101	11010000	01100011
22	10000110	00100100	00011000
23	00000101	11100100	01111011
24	00000110	00100110	01010101

6 Контрольные вопросы

- 1) Какие помехоустойчивые коды называют блоковыми?
- 2) Что такое расстояние Хэмминга?
- 3) Что называют кодовым расстоянием?
- 4) Какие коды называют линейными блоковыми?
- 5) Как определяется кодовое расстояние для линейного блокового кода?
- 6) Как с помощью порождающей матрицы линейного блокового кода осуществляется кодирование информационных слов?
- 7) Что такое совершенные коды?
- 8) Какими соображениями руководствуются при построении матрицы-дополнения для порождающей матрицы линейного блокового кода?
- 9) Как с помощью проверочной матрицы линейного блокового кода можно определить принадлежность кодового вектора данному коду?
- 10) Что понимают под синдромом при декодировании линейных блоковых кодов?
- 11) Каким образом строится стандартная таблица декодирования линейного блокового кода

Практическая работа №8 «Построение кода Хэмминга для обнаружения и исправления одиночных ошибок»

1 Цель работы

- приобрести умение строить код Хэмминга для обнаружения и исправления одиночных ошибок в кодовых словах.

2 Основные теоретические сведения

Систематические коды представляют собой блочные корректирующие коды, в которых информационные и проверочные символы расположены по строго определенной системе и всегда занимают строго определенные места в кодовых словах.

Наиболее известными систематическими кодами, получившими широкое практическое применение, являются коды *Хэмминга*.

Двоичный код Хэмминга строится следующим образом:

1) Определяется число k проверочных символов из условия: $2^k > m + k + 1$; где m - число информационных символов.

2) Выбираются места расположения проверочных символов из условия, чтобы проверочные символы участвовали только в одной операции подсчета четности с целью упрощения процесса кодирования. Такими местами являются символы с номерами, являющиеся целыми степенями числа 2, т. е. 1, 2, 4, 8, 16 и т. д. Символы в кодовых словах Хэмминга нумеруются слева направо.

3) Определяются значения символов слова, называемого *синдромом*:

$$S_k S_{k \cdot h} \dots S_2 S_h$$

из уравнений

$$S_1 = x_1 \oplus x_2 \oplus x_5 \oplus x_7 \oplus \dots = 0,$$

т.е. складываются по модулю 2 значения тех символов, двоичное представление номеров которых содержат 1 в последнем разряде:

$$S_2 = x_2 \oplus x_3 \oplus x_6 \oplus x_7 \oplus \dots = 0,$$

т.е. складываются значения тех символов, двоичное представление номеров которых содержат 1 в предпоследнем

разряде.

Аналогично получаются выражения для нахождения значений S_3, S_4, \dots, S_k .

Так:

$$S_3 = x_4 \oplus x_5 \oplus x_6 \oplus x_7 \oplus \dots = 0;$$

$$S_4 = x_8 \oplus x_9 \oplus x_{10} \oplus x_{11} \oplus \dots = 0;$$

где $x_1, x_2, x_3, x_4, \dots$ - значения символов с номерами 1, 2, 3, 4, ...

Местоположение ошибки, то есть определение символа с ошибкой осуществляется по значению синдрома. Если синдром состоит одних нулей, т. е.

$$S_k S_{k-1} \dots S_2 S_1 = 00 \dots 00,$$

то ошибка отсутствует. Если в синдроме есть символы, отличные от 0, то это говорит о наличии ошибки. Например, если $S_4 S_3 S_2 S_1 = 1000$, то это означает, что ошибка содержится в восьмом символе, так как $8_{10} = 1000_2$.

Другими словами, синдром в коде Хэмминга определяет номер символа с ошибкой. Исправление ошибки осуществляется заменой 0 на 1 либо наоборот 1 на 0.

Для исправления одиночной и обнаружения двойной ошибки, кроме проверок по синдрому, следует проводить еще одну проверку на четность для каждого кодового слова Хэмминга. Чтобы осуществить такую проверку, следует в конце каждого кодового слова добавить проверочный символ таким образом, чтобы сумма единиц в полученном слове всегда была четной. Тогда в случае одной ошибки проверка по синдрому укажет номер ошибочного символа, а проверка на четность укажет наличие ошибки. Если проверка по синдрому укажет на наличие ошибки, а проверка на четность не фиксирует ошибку, то в кодовом слове присутствуют две ошибки.

3 Порядок выполнения работы

Данная практическая работа предполагает выполнение следующих этапов:

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) Для заданных кодовых слов построить кодовые слова

Хэмминга.

4) Осуществить проверку работоспособности кода путем изменения значения одного из символов в любом кодовом слове Хэмминга.

5) Оформить и защитить отчет по практической работе.

4 Пример выполнения работы

Пусть требуется получить кодовое слово Хэмминга для двоичного кодового слова $\bar{x} = 011111$.

Рассматриваемое кодовое слово содержит шесть информационных символов, то есть $m = 6$. Число проверочных символов определяется из условия $2^k > m + k + 1$. При этом минимальное число проверочных символов k , при котором выполняется данное условие, будет $k = 4$, так как $2^4 = 16 > 6 + 4 + 1 = 11$.

Следовательно, кодовое слово Хэмминга для слова $x = 011111$ будет содержать 10 символов, то есть $n = 10$. При этом проверочными символами будут 1, 2, 4, 8, а информационными символами соответственно будут 3, 5, 6, 7, 9, 10.

Пронумеруем и запишем значения информационных символов.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
k_1	k_2	m_1	k_3	m_2	m_3	m_4	k_4	m_5	m_6
		0		1	1	1		1	1

Определим значения проверочных символов. Значение символа $k_1 = x_1$ определяется из условия:

$$S_1 = x_1 \oplus x_3 \oplus x_5 \oplus x_7 \oplus x_9 = 0;$$

$$S_1 = x_1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 = x_1 \oplus 1 = 0.$$

Отсюда $k_1 = x_1 = 1$.

Значения символа $k_2 = x_2$ определяется из условия:

$$S_2 = x_2 \oplus x_3 \oplus x_6 \oplus x_7 \oplus x_{10} = 0;$$

$$S_2 = x_2 \oplus 0 \oplus 1 \oplus 1 \oplus 1 = x_2 \oplus 1 = 0.$$

Отсюда $k_2 = x_2 = 1$.

Значения символа $k_3 = x_4$ определяется из условия:

$$S_3 = x_4 \oplus x_5 \oplus x_6 \oplus x_7 = 0;$$

$$S_3 = x_4 \oplus 1 \oplus 1 \oplus 1 = x_4 \oplus 1 = 0.$$

Отсюда $k_3 = x_4 = 1$.

Значение символа $k_4 = x_8$ определяется из условия:

$$S_4 = x_8 \oplus x_9 \oplus x_{10} = 0;$$

$$S_4 = x_8 \oplus 1 \oplus 1 = x_8 = 0.$$

Отсюда $k_4 = x_8 = 0$.

В итоге кодовое слово Хэмминга для кодового слова $\bar{x} = 011111$ будет иметь вид $\bar{x}_H = 1101111011$.

Проверим работу кода Хэмминга, заменив в полученном кодовом слове значение одного из символов, то есть введем одиночную ошибку в кодовое слово Хэмминга. Например, заменим значение пятого символа слова с 1 на 0, то есть вместо $x_H = 1101111011$ примем $x_H = 1101011011$.

Определим значения символов синдрома S_1, S_2, S_3, S_4 . Для этого пронумеруем символы кодового слова Хэмминга:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	1	0	1	0	1	1	0	1	1

Отсюда значения символов синдрома будут:

$$S_1 = x_1 \oplus x_3 \oplus x_5 \oplus x_7 \oplus x_9 = 1 \oplus 0 \oplus 0 \oplus 1 \oplus 1 = 1;$$

$$S_2 = x_2 \oplus x_3 \oplus x_5 \oplus x_7 \oplus x_{10} = 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1 = 0;$$

$$S_3 = x_2 \oplus x_3 \oplus x_6 \oplus x_7 = 1 \oplus 0 \oplus 1 \oplus 1 = 1;$$

$$S_4 = x_8 \oplus x_9 \oplus x_{10} = 0 \oplus 1 \oplus 1 = 0.$$

Таким образом, синдром имеет следующий вид:

$$S_4 S_3 S_2 S_1 = 0101.$$

Поскольку $0101_2 = 5_{10}$, то ошибка содержится в пятом символе, что и требовалось проверить.

5 Варианты заданий

Пусть требуется получить кодовые векторы Хэмминга для кодовых слов, приведенных в таблице 1. Номер варианта определяется как порядковый номер студента в журнале преподавателя.

Таблица 1 – Варианты заданий для получения кодовых слов

Хэмминга

Вариант	Передаваемые кодовые слова		
1	0111010	1101010	0101011
2	0001010	0001101	0010001
3	0010100	0011000	0011101
4	0011110	0110000	0100010
5	1100011	0001111	0011101
6	1001010	0011001	0100010
7	0001011	0000110	1010001
8	0010101	1000111	1111001
9	1011110	1011000	0100001
10	1000010	1001111	1011101
11	1000011	1100000	0011101
12	0011010	1101010	0101010
13	0000111	1000110	1001000
14	0001110	1001110	0010011
15	0011100	0011010	1011101
16	0010110	0101000	1100010
17	0001011	1101110	0001000
18	1011010	1011001	1100010
19	0001001	0000110	1110001
20	0010111	0100000	1101000
21	1010110	1010000	0110001
22	1000011	0100100	0001000
23	0000010	1100100	0111101
24	0000011	0100110	1010101

6 Контрольные вопросы

1. Какие коды называют систематическими?
2. Сколько ошибок способен обнаруживать и исправлять код Хэмминга?
3. По какому правилу определяют число проверочных символов в коде Хэмминга?
4. В каких местах располагаются проверочные символы в кодовых словах Хэмминга?
5. По какому правилу строятся уравнения для нахождения

проверочных символов в коде Хэмминга?

6. Какую информацию при декодировании кода Хэмминга дает синдром?

7. Каким образом обеспечивается обнаружение двойных ошибок в кодовых словах Хэмминга?

Практическая работа №9

«Определение пропускной способности дискретного канала связи с помехами»

1 Цель работы

- приобрести умение рассчитывать пропускную способность дискретного симметричного канала связи при наличии помех.

2 Основные теоретические сведения

Под **каналом связи** подразумевается совокупность средств, предназначенных для передачи информации от данного источника сообщений к адресату.

Если сигналы на входе и выходе канала представляют собой последовательности символов, составленные из алфавитов A и B , то такой канал называется **дискретным**.

Наличие помех в канале связи приводит к тому, что посланный сигнал с вероятностями, зависящими от самого канала, может переходить в различные выходные сигналы.

Модель дискретного канала задана, если для любых последовательностей передаваемых символов указано правило вычисления условной вероятности $p(a|b)$. Условные вероятности канала записывают в виде **канальной матрицы**.

$$P(a|b) = \begin{bmatrix} p(a_1|b_1) p(a_1|b_2) \dots p(a_1|b_n) \\ p(a_2|b_1) p(a_2|b_2) \dots p(a_2|b_n) \\ \dots \quad \dots \quad \dots \quad \dots \\ p(a_n|b_1) p(a_n|b_2) \dots p(a_n|b_n) \end{bmatrix}$$

В условиях отсутствия помех **скорость передачи информации** по каналу связи определяется количеством информации, переносимым символом сообщения в единицу времени, и равна:

$$V = nH, \text{ бит/сек.};$$

где n - количество символов, вырабатываемых источником сообщений за единицу времени; H - энтропия источника сообщений.

Скорость передачи информации также может быть представлена как

$$V = H / \tau, \text{ бит/сек.};$$

где τ - время передачи одного двоичного символа.

Таким образом, скорость передачи информации определяется относительно первичного алфавита и зависит от его энтропии.

Пропускная способность канала связи - это максимальная скорость передачи информации по данному каналу.

При отсутствии помех выражение для пропускной способности отличается от выражения для скорости тем, что пропускную способность характеризует максимальная энтропия:

$$C = nH_{\max} = n \cdot \log_2 q, \text{ бит/сек.};$$

где q - основание кода.

При наличии помех пропускная способность канала связи определяется следующим образом:

$$C = n[H(A) - H(A|B)] = n[H(B) - H(B|A)], \text{ бит/сек.};$$

где $H(A)$, $H(B)$ - энтропии источника и получателя сообщений;
 $H(A|B)$, $H(B|A)$ - условные энтропии источника сообщений относительно получателя и получателя относительно источника.

3 Порядок выполнения работы

- 1) Изучить методические указания к практической работе.
- 2) Пройти собеседование с преподавателем и получить задание для выполнения работы.
- 3) Определить значения совместных вероятностей и построить матрицу совместных вероятностей для объединенной системы.
- 4) Определить условные вероятности вида $p(a|b)$ и построить соответствующую матрицу условных вероятностей.
- 5) Найти безусловные энтропии источника и получателя сообщений, а также условные энтропии.
- 6) Определить пропускную способность дискретного канала связи.
- 7) Оформить и защитить отчет по выполнению практической работы.

4 Пример выполнения работы

Пусть требуется определить пропускную способность канала связи для двух систем A (источник) и B (получатель), если известны вероятности появления символов в сообщениях на выходе системы A : $p(a_1) = 0,1$; $p(a_2) = 0,4$; $p(a_3) = 0,5$ и матрица условных вероятностей:

$$P(b|a) = \begin{bmatrix} 1 & 0 & 0 \\ 0.25 & 0.75 & 0 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

Также известно, что каждый символ сообщений вырабатывается за 0,01 сек.

Решение

Найдем значения совместных вероятностей и построим матрицу совместных вероятностей для объединенной системы:

$$p(a_1, b_1) = p(a_1) \cdot p(b_1|a_1) = 0.1 \cdot 1 = 0.1;$$

$$p(a_2, b_1) = p(a_2) \cdot p(b_1|a_2) = 0.4 \cdot 0.25 = 0.1;$$

$$p(a_3, b_1) = p(a_3) \cdot p(b_1|a_3) = 0.5 \cdot 0 = 0;$$

$$p(a_1, b_2) = p(a_1) \cdot p(b_2|a_1) = 0.1 \cdot 0 = 0;$$

$$\bullet p(a_2, b_2) = p(a_2) \cdot p(b_2|a_2) = 0.4 \cdot 0.75 = 0.3;$$

$$p(a_3, b_2) = p(a_3) \cdot p(b_2|a_3) = 0.5 \cdot 0.2 = 0.1;$$

$$p(a_1, b_3) = p(a_1) \cdot p(b_3|a_1) = 0.1 \cdot 0 = 0;$$

$$p(a_2, b_3) = p(a_2) \cdot p(b_3|a_2) = 0.4 \cdot 0 = 0;$$

$$p(a_3, b_3) = p(a_3) \cdot p(b_3|a_3) = 0.5 \cdot 0.8 = 0.4;$$

$$P(a,b) = \begin{bmatrix} 0.1 & 0 & 0 \\ 0.1 & 0.3 & 0 \\ 0 & 0.1 & 0.4 \end{bmatrix}$$

Найдем вероятности появления сигналов на входе системы B суммированием столбцов полученной матрицы:

$$p(b_1) = 0.1 + 0.1 = 0.2;$$

$$p(b_2) = 0.3 + 0.1 = 0.4;$$

$$p(b_3) = 0.4.$$

$$\sum_i p(b_i) = 0.2 + 0.4 + 0.4 = 1.$$

Найдем условные вероятности вида $p(b|a)$ и построим соответствующую матрицу условных вероятностей:

$$p(a_1, b_1) = p(a_1, b_1) / p(b_1) = 0.1 / 0.2 = 0.5;$$

$$p(a_2, b_1) = p(a_1, b_2) / p(b_2) = 0 / 0.4 = 0;$$

$$p(a_3, b_1) = p(a_1, b_3) / p(b_3) = 0 / 0.4 = 0;$$

$$p(a_1, b_2) = p(a_2, b_1) / p(b_1) = 0.1 / 0.2 = 0.5;$$

$$p(a_2, b_2) = p(a_2, b_2) / p(b_2) = 0.3 / 0.4 = 0.75;$$

$$p(a_3, b_2) = p(a_2, b_3) / p(b_3) = 0 / 0.4 = 0;$$

$$p(a_1, b_3) = p(a_3, b_1) / p(b_1) = 0 / 0.2 = 0;$$

$$p(a_2, b_3) = p(a_3, b_2) / p(b_2) = 0.1 / 0.4 = 0;$$

$$p(a_3, b_3) = p(a_3, b_3) / p(b_3) = 0.4 / 0.4 = 1;$$

$$P(a, b) = \begin{bmatrix} 0.5 & 0 & 0 \\ 0.5 & 0.75 & 0 \\ 0 & 0.25 & 1 \end{bmatrix}$$

Отсюда безусловные энтропии источника и получателя сообщений будут:

$$H(A) = - (0.1 \log_2 0.1 + 0.4 \log_2 0.4 + 0.5 \log_2 0.5) = 0.332 + 0.528 + 0.5 = 1.36 \text{ бит/символ};$$

$$H(B) = - (0.2 \log_2 0.2 + 0.4 \log_2 0.4 + 0.4 \log_2 0.4) = 0.464 + 0.528 + 0.528 = 1.520 \text{ бит/символ}.$$

Условные энтропии будут:

$$H(A|B) = - \sum_i \sum_j p(b_j) p(a_i|b_j) \log_2 p(a_i|b_j) = - [0.2 \cdot 0.5 \log_2 0.5 + 0.4(0.75 \log_2 0.75 + 0.25 \log_2 0.25)] = 0.3244 + 0.2 = 0.5244 \text{ бит/символ};$$

$$H(B|A) = - \sum_i \sum_j p(a_i) p(b_j|a_i) \log_2 p(b_j|a_i) = - [0.1 \cdot 1 \cdot \log_2 1 + 0.4(0.75 \log_2 0.75 + 0.25 \log_2 0.25) + 0.5(0.84 \log_2 0.8 + 0.24 \log_2 0.2)] = 0.3244 + 0.2 = 0.3244 + 0.3610 = 0.6854 \text{ Бит/символ}.$$

Пропускная способность канала связи будет:

$$C = n[H(A) - H(A|B)] = 100(1.36 - 0.524) \approx 8.3 \text{ бит/сек}; \text{ или}$$

$$C = n[H(B) - H(B|A)] = 100(1.52 - 0.685) \approx 8.3 \text{ бит/сек}.$$

5 Варианты заданий

Номер варианта определяется как порядковый номер студента в

журнале преподавателя

Таблица 1 – Вероятности появления символов в сообщениях на выходе источника

Вариант	Символы алфавита				Вариант	Символы алфавита			
	a_1	a_2	a_3	a_4		a_1	a_2	a_3	a_4
1	0,23	0,27	0,39	0,11	13	0,24	0,16	0,28	0,32
2	0,21	0,32	0,09	0,38	14	0,15	0,41	0,35	0,09
3	0,18	0,42	0,25	0,15	15	0,25	0,35	0,09	0,31
4	0,08	0,26	0,32	0,34	16	0,40	0,17	0,20	0,23
5	0,37	0,13	0,09	0,41	17	0,40	0,10	0,17	0,33
6	0,16	0,31	0,24	0,29	18	0,36	0,08	0,14	0,42
7	0,10	0,40	0,31	0,19	19	0,19	0,31	0,08	0,42
8	0,24	0,27	0,16	0,33	20	0,34	0,18	0,26	0,22
9	0,15	0,25	0,42	0,18	21	0,31	0,19	0,41	0,09
10	0,17	0,42	0,33	0,08	22	0,37	0,15	0,23	0,25
11	0,40	0,30	0,08	0,22	23	0,22	0,08	0,29	0,41
12	0,26	0,31	0,24	0,19	24	0,12	0,30	0,18	0,40

Таблица 2 – Матрицы переходных вероятностей $P(b|a)$

Вар.	Пер. симв.	Принятые символы				Вар.	Пер. симв.	Принятые символы			
		b_1	b_2	b_3	b_4			b_1	b_2	b_3	b_4
1	a_1	0,95	0,03	0,02	0	4,	a_3	0,98	0,02	0	0
7	a_2	0,02	0,97	0,01	0	ю,	a_2	0,03	0,96	0,01	0
13	a_3	0	0	0,99	0,01	16,	a_3	0	0,02	0,94	0,04
19	a_4	0	0,02	0,04	0,94	22	a_4	0	0,01	0,02	0,97
2	a_5	0,97	0,02	0,01	0	5,	a_1	0,94	0,03	0,02	0,01
8	a_2	0,02	0,98	0	0	П,	a_2	0,01	0,99	0	0
14	a_3	0	0,02	0,95	0,03	17,	a_3	0	0,03	0,96	0,01
20	a_4	0	0	0,01	0,99	23	a_4	0	0,01	0,04	0,95
3	a_1	0,96	0,03	0,01	0	6,	a_3	0,99	0,01	0	0
9	a_2	0,03	0,95	0,02	0	12,	a_2	0,02	0,94	0,03	0,01
15	a_3	0	0,01	0,97	0,02	18,	a_3	0	0,02	0,98	0
21	a_4	0	0	0,02	0,98	24	a_4	0	0,01	0,03	0,96

6 Контрольные вопросы

- 1) Что понимают под каналом связи?
2. Какие каналы связи называют дискретными?
3. Каким образом задают описание дискретного канала связи с

помехами?

4. Как определяют скорость передачи информации по дискретному каналу связи?

5. Что такое пропускная способность канала связи?

6. Как определяется пропускная способность дискретного канала при отсутствии помех?

7. Как определяется пропускная способность дискретного канала с помехами?

Список использованных источников

1. Хмелевская, А.В. Основы теории информации и кодирования [Текст]: учеб. пособие / А.В. Хмелевская, А.М. Потапенко; Юго-Зап.гос. ун-т. - Курск, 2016. - 220 с. Библиогр.: с.210
2. Гульятеева, Т. А. Основы теории информации и криптографии [Электронный ресурс]: конспект лекций / Т. А. Гульятеева. - Новосибирск: НГТУ, 2010. - 88 с. - ISBN 978-5-7782-1425-5: Б. ц.
3. Белов, В. М. Теория информации [Текст]: учебное пособие для студентов, обучающихся по направлению подготовки 11.03.02 «Инфокоммуникационные технологии и системы связи» / В. М. Белов, С. Н. Новиков, О. И. Солонская. - Москва: Горячая линия - Телеком, 2017. - 144 с. : - Библиогр.: с. 142-143. - ISBN 978-5-9912-0237-4 Б.ц.
4. Тихонов В. И. Случайные процессы. Примеры и задачи [Текст]: учебное пособие / В. И. Тихонов; под ред. В. В. Сизых. - М.: Горячая линия - Телеком, 2012 - .Т. 5 : Оценка сигналов, их параметров и спектров. Основы теории информации. – 400 с.
5. Акулиничев, Ю. П. Теория электрической связи [Текст]: учебное пособие / Ю. П. Акулиничев. - СПб.; М.; Краснодар: Лань, 2010. – 240 с.
6. Биккенин, Р. Р. Теория электрической связи [Текст]: учебное пособие / Р. Р. Биккенин, М. Н. Чесноков. - М.: Академия, 2010. – 336 с.
7. Холево, А. С. Квантовые системы, каналы, информация [Текст]: учебное пособие / А. С. Холево. - М.: МЦНМО, 2010. – 328 с.
8. Котоусов А.С. Теория информации. Учебное пособие для вузов. – М.: Радио и связь, 2003.
9. Липкин И.А. Статистическая радиотехника. Теория информации и кодирования. – М.: Вузовская книга, 2002.

