Учреждение образования

«Белорусский государственный технологический университет»

**Отчёт за практические занятия по предмету:**

Защита информации и надежность информационных систем

Выполнила:

Студентка 2 курса 1 группы ФИТ

Гринцевич Юлия Сергеевна

**2022г.**

# **Лабораторная работа №1**

**Разработка политики информационной безопасности консалтинговой компании**

**Цель работы**: Разработать проект политики информационной безопасности консалтинговой компании.

**Введение**

В нынешнее время все компании нуждаются в информационных технологиях для того, чтобы выдерживать конкуренцию в условиях повсеместного использования высокотехнологичных решений. Информационные ресурсы в современном обществе являются наиболее значимыми, но при этом, очень уязвимыми т.к. механизм их распространения становится все более гибким и количество возможностей их передачи растет. В связи с этим информационная безопасность — одно из важнейших условий функционирования любой структуры.

Жизнь современного общества немыслима без информационных технологий. Программные средства обслуживают банковские системы, распределяют энергию, следят за расписанием поездов. Они обеспечивают хранение информации, ее обработку и предоставление потребителям, реализуя таким образом информационные технологии.

Темпы развития современных информационных технологий значительно опережают темпы разработки рекомендательной и нормативно-правовой базы руководящих документов, действующих на территории Беларуси. Поэтому решение вопроса о разработке эффективной политики информационной безопасности на современном предприятии обязательно связано с проблемой выбора критериев и показателей защищенности, а также эффективности корпоративной системы защиты информации.

Однако именно высокая степень автоматизации порождает риск снижения личной, информационной и государственной безопасности, делая ее чрезвычайно уязвимой.

Вследствие этого, в дополнение к требованиям и рекомендациям стандартов, Конституции, законам и иным руководящим документам приходится использовать ряд международных рекомендаций. В том числе адаптировать к отечественным условиям и применять на практике методики международных стандартов, а также использовать методики управления информационными рисками в совокупности с оценками экономической эффективности инвестиций в обеспечение защиты информации предприятия.

**Обоснование актуальности**

Любой компании, ведущей свой бизнес, приходится работать с информацией везде и всегда, как при планировании долгосрочных стратегических задач, так и при рутинных канцелярских делах. В процессе бурного развития науки и техники появилось немало методов и средств работы с информацией, порой настолько сложных, что применение их, их адаптация к специфике конкретной компании и условий, в которых она находится, сейчас уже является отдельным классом задач, решаемых специализированными организациями.

Каждая фирма имеет дело с внешними и внутренними источниками информации, с выполнением таких операций, как сбор информации, ее первичная обработка, отображение, кодирование и декодирование, преобразование, запоминание (хранение), передача, защита. В средних и тем более крупных по размеру компаниях осуществление вышеуказанных операций немыслимо без применения информационных технологий. Под информационными технологиями здесь понимаются: обычные средства; средства на основе вычислительной техники; методы связи.

В век высоких технологий компаниям просто необходимо использовать информационных ресурсы и технологии. Согласно исследованиям, многие компании практически не защищены от кражи и утери конфиденциальных данных. Сейчас, как никогда, тема информационной безопасности актуальна для компаний.

Новые технологии позволяют собрать клиентскую базу, анализировать и структурировать ее, а также изучать информацию, создавать базы стандартных ответов в типичных ситуациях.

В числе наиболее важных мер по обеспечению информационной безопасности является аудит информационной безопасности, создание документа, регулирующего процесс обеспечения информационной безопасности, настройку систем защиты, проведения тренинга среди сотрудников, страхование рисков во всех сферах.

**Общие положения**

Цели и задачи

Концептуальная схема информационной безопасности компании направлена на защиту его информационных активов от угроз, исходящих от противоправных действий злоумышленников, уменьшение рисков и снижение потенциального вреда от аварий, непреднамеренных ошибочных действий персонала, технических сбоев, неправильных технологических и организационных решений в процессах обработки, передачи и хранения информации и обеспечение нормального функционирования технологических процессов.

- разработка и совершенствование нормативно-правовой базы обеспечения информационной безопасности;

- выявление, оценка и прогнозирование угроз информационной безопасности;

- организация технической защиты информации, участие в проектировании систем защиты;

- проведение периодического контроля состояния ИБ, учет и анализ результатов с выработкой решений по устранению уязвимостей и нарушений;

- контроль за использованием закрытых каналов связи и ключей с цифровыми подписями;

- организация плановых проверок режима защиты, и разработка соответствующей документации, анализ результатов, расследование нарушений;

- разработка и осуществление мероприятий по защите персональных данных;

- организация взаимодействия со всеми структурами, участвующими в их обработке, выполнение требований законодательства к информационным системам персональных данных, контроль действий операторов, отвечающих за их обработку.

**Объекты защиты**

Под организационной структурой управления понимается состав отделов, служб и подразделений в аппарате управления, системная их организация, характер соподчиненности и подчиненности друг другу и высшему органу управления, а также набор координационных и информационных связей, порядок распределения функций управления по различным уровням и подразделениям.

Консультанты по управлению большую часть времени работают в организациях клиентов во временно сформированной команде. После завершения консалтингового проекта команда расформировывается и следующий проект выполняет вновь сформированная команда с другим составом консультантов. Для стабильности внутренней организации в компании создаются постоянные подразделения, к которым прикрепляют консультантов. Организационно консалтинговые компании, как правило, формируются по дивизиональному принципу.

**Функциональные подразделения**. Такие подразделения формируются по основным функциям управления, например, общее руководство, финансы, маркетинг, кадры, стратегия и т.д.

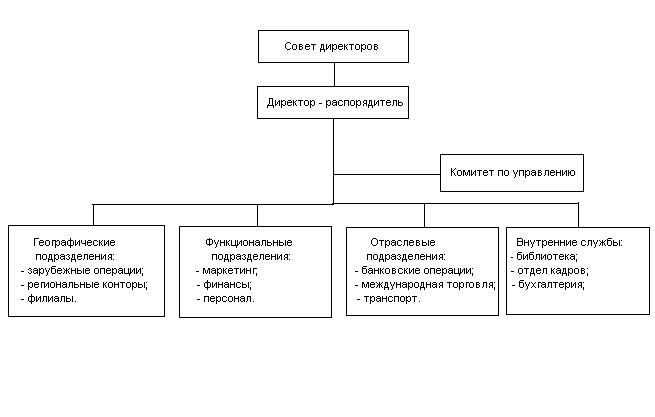
**Отраслевые подразделения**(например, по строительству, банковскому делу, международной торговле, транспорту, здравоохранению и т.д.) создаются в том случае, если имеется достаточно высокий спрос на такие виды услуг.

**Географические (территориальные) подразделения** создаются, чтобы приблизиться к клиентам и снизить транспортные, почтовые и другие расходы.

Такие подразделения формируются в виде региональной конторы или филиала.

Региональная контораявляется малым подразделением и создается с целью маркетинга и связи с клиентами в ограниченных географических областях. Такие подразделения укомплектовываются несколькими всесторонне образованными консультантами по управлению, которые выполняют консалтинговые проекты совместно с консультантами из головной компании.

Филиалыявляются полностью укомплектованными подразделениями и в состоянии справиться с выполнением большинства проектов силами собственного персонала. Они эффективны, если объемы работ в данной области относительно стабильны



Тип высшего руководства консалтинговой компании зависит от ее размеров и правовой формы организации бизнеса. В корпорациях это совет директоров. В мелких компаниях высшим руководителем является директор-распорядитель.

В крупной компании может быть также сформирован комитет по управлению, в состав которого входят члены внешнего совета.

Ключевое положение в управленческой иерархии крупной компании занимает директор-распорядитель (президент, генеральный директор и т.п.). Как правило, это опытный консультант, имеющий способности к управлению специфическим коллективом. Некоторые компании приглашают свое высшее руководство со стороны, выбирая его из лиц, которые показали себя хорошими руководителями и не обязательно работали консультантами.

**Основные угрозы и их источники**

Угрозы информационной безопасности могут быть классифицированы по различным признакам:

*– по аспекту информационной безопасности, на который направлены угрозы (конфиденциальности, целостности, доступности)*

*– по степени преднамеренности действий (случайные, преднамеренные)*

*– по расположению источника угроз (внутренние, внешние)*

*– по размерам наносимого ущерба (общие, локальные, частные)*

*– по степени воздействия на информационную систему (пассивные, активные).*

Основными угрозами ИБ являются:

*– несанкционированное вмешательство;*

*– нарушение функционирования;*

*– несанкционированный доступ к информации, хранимой в базах данных и передаваемой по каналам передачи данных.*

В результате реализации угроз ИБ могут быть нарушены:

*– целостность (утрата, уничтожение, модификация) информации;*

*– доступность (блокирование) информации и отдельных элементов;*

*– конфиденциальность информации.*

Обеспечение ИБ осуществляется по следующим направлениям, реализуемым организационно-техническими мерами защиты.

Физическая защита, включая:

– защиту технических средств обработки, хранения и передачи информации;

– защиту сооружений и помещений.

Защита при эксплуатации средств обработки, хранения и передачи информации и использовании информационных ресурсов, включая:

– защиту от вредоносного ПО;

– резервирование серверов, сетевого оборудования, средств защиты и каналов передачи данных;

– обеспечение безопасности сетевой инфраструктуры;

– защиту программного обеспечения;

– регистрацию и учет событий ИБ;

– организацию безопасного использования, преобразования, хранения и передачи информации;

– криптографическую защиту.

Контроль доступа, в том числе:

– управление доступом пользователей;

– определение ответственности пользователей;

– контроль доступа к прикладным системам;

– контроль доступа к ОС;

– обеспечение безопасности при использовании мобильных устройств;

– обеспечение безопасности в беспроводных сетях;

– контроль доступа к сетевому оборудованию.

**Оценка угроз, рисков и уязвимостей**

Согласно промежуточным результатам исследования, все компании имеют незакрытые критические уязвимости на ПК и серверах, половина – на сетевом оборудовании. 15% компаний не имеют даже минимальной системы защиты, 39% – уязвимы для действий инсайдеров, а 62% – не защищены от внешних угроз. Причинами являются отсутствие шифрования данных; реагирования на инциденты; не разграничение прав доступа; из всех средств защиты присутствуют лишь антивирус и слабые пароли.

На практике это, в частности, означает высокую вероятность взлома; моментальную доступность конфиденциальной информации при утере или выемке устройства, на которой она хранится; неотслеживаемость действий инсайдеров; невозможность узнать и отреагировать на кибератаку.

В частности, нарушение конфиденциальности данных компании зачастую приводит к вымогательству и шантажу, инсайдерской торговле и недобросовестной конкуренции. Это не только нанесет урон репутации – фирма понесет ответственность – от финансовой до уголовной.

Цель оценивания рисков состоит в определении характеристик рисков для информационной системы и ее ресурсов. При оценивании рисков учитывается:

* ценность ресурсов;
* оценка значимости угроз;
* эффективность существующих и планируемых средств защиты.

Показатели ресурсов или потенциальное негативное воздействие на деятельность организации можно определять несколькими способами: количественными (например, стоимостными), качественными (могут быть построены на использовании таких понятий, как, умеренный или чрезвычайно опасный) или их комбинацией.

Таблица 1 – Условная численная шкала для оценки ущерба it-компании

|  |  |
| --- | --- |
| Величина ущерба | Описание |
| 1 | Раскрытие данных принесет ничтожно малый моральный и экономический ущерб компании |
| 2 | Ущерб от атаки есть, но основные операции и положения затронуты не будут |
| 3 | Значительные финансовые потери и потери доверия со стороны клиента, ощутимая часть которых уйдет |
| 4 | Потери очень значительные, компания теряет свое положение на рынке на период до 2-ух лет. Для восстановления положения потребуются крупные финансовые займы |
| 5 | Компания прекращает существование |

В таблице 1 показаны последствия различной величины ущерба

Можно конкретизировать определение вероятности наступления угрозы ресурсу. Вопрос о том, как провести границу между допустимыми и недопустимыми рисками, решается пользователем. Очевидно, что разработка ПБ требует учета специфики IT-компании. На таблице 2 приведена шкала вероятности того, что угроза будет реализована.

Таблица 2 – Вероятностно-временная шкала реализации НСД к информационным ресурсам

|  |  |
| --- | --- |
| Вероятность события | Средняя частота события (НСД) |
| 0 | Данный вид атаки отсутствует |
| 0,1 | Реже, чем раз в год |
| 0,2 | Около 1 раза в год |
| 0,3 | Около 1 раза в месяц |
| 0,4 | Около 1 раза в неделю |
| 0,5 | Практически ежедневно |

Также можно создать таблицу рисков, на этапе анализа которой риски задаюются некоторым максимально допустимым уровнем (порогом). Далее проверятся каждая строка таблицы: превышен или непревышен порог для значения риска, связанного с анализируемой атакой.

Таблица 3 – Оценка рисков

|  |  |  |  |
| --- | --- | --- | --- |
| Описание атаки | Ущерб | Вероятность | Риск (Ущерб \* Вероятность) |
| Спам (переполнение почтового ящика) | 1 | 0,4 | 0,4 |
| Копирование жесткого диска из центрального офиса | 3 | 0,1 | 0,3 |
| НСД к корпоративной информации | 4 | 0,2 | 0,8 |
| Мониторинг внутренней работы со стороны специальных служб | 3 | 0,2 | 0,6 |
| Отказы и сбои в работе оборудования | 3 | 0,3 | 0,9 |
| Проникновение на рабочие места не работников отдела, которые имеют туда доступ | 3 | 0,1 | 0,3 |
| Ошибки в работе системных администрторов | 2 | 0,4 | 0,8 |
| Итого | 19 | 1,7 | 4,1 |

Если такое превышение имеет место, данная атака должна рассматриваться с точи зрения одной из первоочередных целей разработки политики безопасности (таблица 3).

**Разработка мер защиты**

Информационная безопасность компаниидолжна обеспечиваться целым комплексом мер, среди которых:

* административно-правовые;
* организационные;
* программно-технические.

Данные меры следует применять совместно. Опираться система защиты должна на управление персоналом компании и контроль над ним. Меры технического характера не менее важны, но не могут существовать в отрыве от организационных мер.

Административные меры безопасности включают в себя разработку внутренних нормативных документов, обеспечивающих информирование сотрудников о системе действий, необходимых для обеспечения информационной безопасности. В компании должно быть организовано ознакомление персонала с внутренними нормативными документами.

Служба безопасности компанииразрабатывает и предлагает на утверждение руководства политику защиты конфиденциальной информации, который содержит:

* основные принципы защиты конфиденциальной информации в компании;
* контроль исполнения требований законодательства и внутренних документов;
* обязанности каждого сотрудника в части защиты доверенных ему сведений;
* меры ответственности за нарушение положений документа;
* задачи руководства по обеспечению охраны информации;
* проведение профилактических мер при приеме на работу и увольнении работников;
* регламенты обращения с компьютерной техникой и средствами коммуникации;
* контроль соответствия бизнес-процессов требованиям политики;
* реагирование на инциденты, локализацию и минимизацию последствий;
* анализ новых рисков ИБ компании;
* определение действий при возникновении чрезвычайных ситуаций.

Все трудовые договоры должны содержать перечень информации, носящей характер коммерческой, а в самих договорах должны быть предусмотрены меры ответственности за ее разглашение.

Организационные меры безопасности направлены на устранение внутренней угрозы утечки информации и мотивацию сотрудников на соблюдение утвержденных регламентов. Эти меры предпринимаются службой безопасности во взаимодействии с сотрудниками служб управления персоналом.

В числе организационных мер обеспечения информационной безопасности можно назвать следующие:

* установление различных степеней допуска сотрудников к сведениям, содержащим коммерческую тайну;
* ограничение круга лиц, имеющих допуск к конфиденциальной информации it-компании;
* организация порядка использования материальных носителей, установление контроля над копированием и сканированием документов, ограничение доступа сотрудников к внешней электронной почте;
* проведение периодических проверок соблюдения регламентов;
* привлечение специалистов для проведения тренингов по защите информации;
* проведение мероприятий по созданию режима коммерческой тайны;
* внесение в договоры компании с клиентами норм, касающихся обязательств соблюдения последними режима коммерческой тайны в отношении переданной им информации;
* привлечение к ответственности лиц, виновных в разглашении информации.

Иногда система работы с безопасностью информации требует создания в компании специального подразделения, в чьи функции будет входить только эта деятельность.

Также следует учитывать, что при проектировании большинства информационных систем уровень защиты от внешнего проникновения был значительно ниже, чем необходим в настоящее время. Среди организационных мер может быть и их аудит, который установит соответствие современным стандартам.

Технические меры рассчитаны на использование действенных технических средств защиты. Для их реализации используются аппаратные, программные и криптографические средства. Первые предполагают установку систем резервного копирования и защиту от несанкционированного проникновения, вторые отвечают за работу антивирусов и иных защитных программ, третьи обеспечивают шифрование всей хранимой и передаваемой по каналам связи информации. Наиболее часто для защиты информации применяются межсетевые экраны и системы обнаружения вторжений. Технические средства требуют постоянного обновления и модернизации, так как скорость устаревания программных продуктов очень высока.

Технические меры безопасности включают (но не ограничены ими):

* регистрацию, проверку пользователя при помощи капчи;
* построение периметра безопасности защищаемых объектов;
* организацию круглосуточной охраны охраняемых объектов, в том числе с использованием технических средств безопасности;
* организацию противопожарной безопасности складов;
* контроль доступа работников компаниив помещения ограниченного доступа;
* использование лицензионного программного обеспечения и сертифицированных средств защиты информации;
* применение комплексной антивирусной защиты;
* использование специальных комплексов ИК (как защита электронных информационных ресурсов, так и защита от утечки по электромагнитным и акустическим каналам);
* обеспечение регулярного резервного копирования информации;
* контроль за правами и действиями пользователей, в первую очередь, привилегированных;
* применение систем криптографической защиты информации;
* обеспечение безотказной работы аппаратных средств;
* мониторинг состояния критичных элементов информационной системы.

Следует провести комплексную оценку с привлечением специализированных людей. Также, эффективной мерой по защите является проведение тестовых экспериментов по атаке на фирму: наем специализированных людей, которые проведут атаку на сетевые ресурсы компании и выявят пробелы в защите. Данная операция безусловно является дорогостоящей, однако впоследствии, это сэкономит средства компании.

Комплексное применение современных технических средств в работе службы безопасности it-компании может обеспечить высокий уровень защиты информации от утечек и несанкционированного доступа.

**Вывод**

Мною были выявлены все возможные угрозы и их источники информационной безопасности в ИТ-компании, а также проанализированы и оценены собранные данные.

На основании этого была разработана концепция, основные элементы политики безопасности для ИТ-компании.

Также были разработаны мероприятия по внедрению разработанной мной политики безопасности на предприятии.

Результаты проделанной работы оформлены в виде описания разработанной политики безопасности а также плана мероприятий по ее реализации.

Обобщая все вышесказанное и учитывая возможные виды угрозы/атаки, можем выделить следующие рекомендации, следуя которым возможно снизить риски опасного воздействия и их последствий:

* четкая и строгая иерархия должностей и полномочий в компании. Каждый должен заниматься строго своим заданием;
* обязательная экстренная связь офиса с милицией и пожарной службой (наличие кнопок экстренного вызова) и четкий инструктаж персонала на случай чрезвычайного происшествия;
* строгий подбор сотрудников с привлечением, при необходимости, баз данных милиции;
* защита серверов компании и важной корпоративной почты;
* использование новейших средств защиты (антивирусные продукты, файерволы) персональных компьютеров сотрудников и обязательное использование лишь лицензионных продуктов;
* разграничение доступа к финансовым отделам.
* проведение регулярных бесед и инструктажей с сотрудниками;
* наличие в офисе наглядного отображения плана по работе во время чрезвычайных ситуаций.

Политика безопасности должна быть разработана так, чтобы её взлом стоил дороже, чем сама информация.

**Ответы на вопросы:**

1. *Охарактеризовать актуальность и основные причины проблемы информационной безопасности организации, страны.*

Обеспечение ИБ на предприятиях и в учреждениях, как правило, является неотъемлемой частью общей системы управления, необходимой для достижения уставных целей и задач. Значимость систематической целенаправленной деятельности по обеспечению ИБ становится тем более высокой, чем выше степень автоматизации бизнес-процессов. Значимость обеспечения ИБ в некоторых случаях может определяться наличием в общей системе информационных потоков предприятия сведений, составляющих не только коммерческую, но и государственную тайну, а также другие виды конфиденциальной информации: сведения, составляющие банковскую тайну, различные виды персональных данных, в том числе врачебная тайна, интеллектуальная собственность компаний-партнеров и т. п.

1. *Сформулировать цели и задачи политики информационной безопасности.*

* конфиденциальность данных – доступ есть только у лиц, имеющих на это полномочия;
* доступность информационных систем с находящимися в них данными конкретным пользователям, у которых есть право доступа к таким сведениям;
* целостность данных предполагает блокировку несанкционированного изменения информации;
* подлинность – полнота и общая точность информации;
* неотказуемость – возможность определить источник или авторство информации.

1. *Как правильно проводить оценку рисков?*

Оценка рисков включает в себя следующие **действия и мероприятия**:

* идентификация значимых угроз и уязвимостей для идентифицированных ресурсов;
* оценка вероятности возникновения угроз и уязвимостей;
* вычисление рисков; оценивание рисков по заранее определенной шкале риска.

При оценивании рисков учитываются:

* ценность ресурсов;
* оценка значимости угроз;
* эффективность существующих и планируемых средств защиты.

1. *Что должна включать в себя программа внедрения политики информационной безопасности?*

Политика информационной безопасности должна рассматриваться как система, как комплекс инструментов по защите информации.

Основу таких мероприятий составляют инструкции, содержащие подробное описание (алгоритмы) действий по организации информационной защиты и обеспечению разработанных стандартов и процедур, и план мероприятий по обучению персонала и тестированию знаний сотрудников, имеющих доступ к информационным ресурсам.

Можно выделить следующие общие **направления мероприятий**:

* *управление персоналом* (для каждой должности должны существовать квалификационные требования по ИБ, в должностные инструкции должны входить разделы, касающиеся информационной безопасности, каждого работника нужно научить мерам безопасности теоретически и на практике);
* *физическая защита инфраструктуры ИВС* (защита от утечки информации по техническим каналам, инженерные способы защиты и т. д.);
* *поддержание работоспособности ИВС* (создание инфраструктуры, включающий в себя как технические, так и процедурные регуляторы и способной обеспечить любой наперед заданный уровень работоспособности на всем протяжении жизненного цикла информационной системы.);
* *реагирование на нарушения режима безопасности ИВС* (может быть регламентировано в рамках отдельно взятой организации. В настоящее время осуществляется только мониторинг компьютерных преступлений в национальном масштабе и на мировом уровне.);
* *планирование восстановительных работ* (слаженность действий персонала во время и после аварии, наличие заранее подготовленных резервных производственных площадок, официально утвержденную схему переноса на резервные площадки основных информационных ресурсов, схему возвращения к нормальному режиму работы).

# **Лабораторная работа №2**

**Элементы теории информации. Параметры и характеристики дискретных информационных систем**

**Цель**: приобретение практических навыков расчета и анализа параметров иы информативных характеристик дискретных ИС.

**2.1 Теоретические сведения**

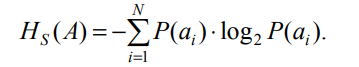
Передача информации (данных) осуществляется между двумя абонентами, называемыми источником сообщения (ИcС) и получателем сообщения (ПС). Третьим элементом информационной системы является канал (среда) передачи, связывающий ИсС и ПС. Отметим также, что и в системах с хранением информации всегда можно выделить ИcС и ПС. В данном случае каналом передачи здесь выступает устройство хранения информации (память). Например, при записи данных в ОЗУ (оперативное запоминающее устройство) компьютера в качестве ИcС и ПС может выступать процессор (соответственно при записи и чтении данных). Таким образом, простейшая информационная система состоит из трех элементов: источника сообщения, канала передачи сообщения и получателя сообщения. Отображение сообщения обеспечивается изменением какойлибо физической величины, характеризующей процесс (например, амплитуда, частота, фаза). Эта величина является информационным параметром сигнала (в общем случае – информационной системы).

Сигналы, как и сообщения, могут быть непрерывными и дискретными. Информационный параметр непрерывного сигнала с течением времени может принимать любые мгновенные значения в определенных пределах. Непрерывный сигнал часто называют аналоговым, а каналы и устройства, функционирующие на основе такого типа сигналов, – аналоговыми. Дискретный сигнал (устройство или канал передачи) характеризуется конечным числом значений информационного параметра. Дискретные сообщения состоят из последовательности дискретных знаков. Часто этот параметр принимает всего два значения (0 или 1). Сообщение или канал его передачи на основе этих двух значений сигнала называют двоичным или бинарным. Построение сигнала по определенным правилам, обеспечивающим соответствие между сообщением и сигналом, называют кодированием. Кодирование в широком смысле – преобразование сообщения в сигнал. Кодирование в узком смысле – представление исходных знаков, называемых символами, в другом алфавите с меньшим числом знаков. Оно осуществляется с целью повышения надежности и преобразования сигналов к виду, удобному для передачи по каналам связи. Последний тип кодирования относится к так называемой прикладной теории кодирования информации, занимающейся поиском и реализацией методов и средств обнаружения несоответствий (ошибок) между переданным Xk и принятым Yk сообщениями.

Рассмотрим основные характеристики и параметры двоичных систем. Важнейшая характеристика источника, получателя или канала – алфавит. Алфавит, А – это общее число знаков или символов (N), используемых для генерации или передачи сообщений. Символы алфавита будем обозначать через {аi}, где 1 ≤ i ≤ N; N – мощность алфавита. Минимальное число знак представляет собой элементарное сообщение, последовательность знаков – сообщение. Набор элементов алфавита, создаваемых дискретным источником сообщений, заранее, априори (до опыта) известен получателю. Элементы теории информации. Параметры и характеристики дискретных систем 23. ИсС в каждый дискретный момент времени выдает один элемент алфавита. Этот элемент сообщения является одним из символов алфавита. Понятно, что ПС заранее не известно, какой это элемент. Если обозначить вероятность выбора каждого элемента алфавита p(аi), то 

Вероятности p(аi) могут быть получены в результате анализа частотных свойств символов алфавита, если на входе такого анализатора принять документ на основе соответствующего алфавита. Причем объем документа должен быть таким, чтобы от частости (частоты) появления каждого символа в анализируемом документе можно было перейти к вероятности соответствующего события.

Энтропию алфавита А = {ai} по К. Шеннону рассчитывают по следующей формуле:



С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита. Частным случаем энтропии Шеннона считается энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита. С учетом этого формулу (2.1) можно преобразовать к виду:



Сообщение Хk, которое состоит из k символов, должно характеризоваться определенным количеством информации I(Хk):



**2.2 Практическое задание**

а) рассчитать энтропию указанных преподавателем алфавитов: один – на латинице, другой – на кириллице (по формуле (2.1) перейти от частоты появления каждого символа алфавита к соответствующей вероятности); в качестве входного может быть принят произвольный электронный текстовый документ на основе соответствующего алфавита; частоты появления символов алфавитов оформить в виде гистограмм (можно воспользоваться приложением MS Excel);

б) для входных документов, представленных в бинарных кодах, определить энтропию бинарного алфавита;

в) используя значения энтропии алфавитов, полученных в пунктах (а) и (б), подсчитать количество информации в сообщении, состоящем из собственных фамилии, имени и отчества (на основе исходного алфавита – (а) и в кодах ASCII – (б)); объяснить полученный результат;

г) выполнить задание пункта (в) при условии, что вероятность ошибочной передачи единичного бита сообщения составляет: 0,1;0,5; 1,0

**2.3 Выполнение практического задания**

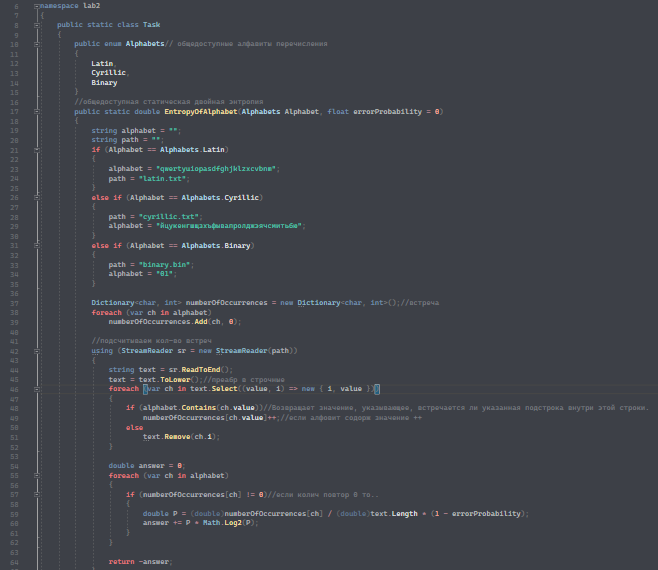


Рис. *1* – Код файла Task.cs

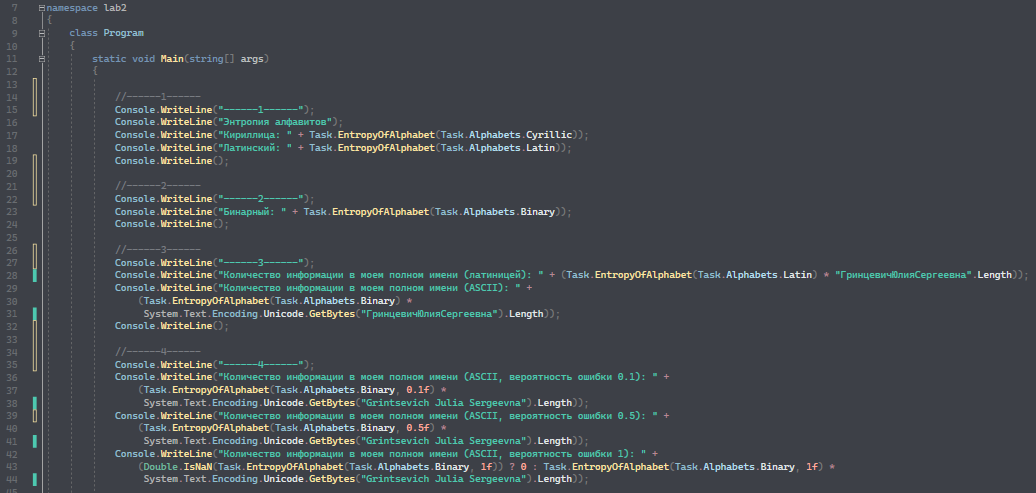


Рис. *2* – Код файла Program.cs

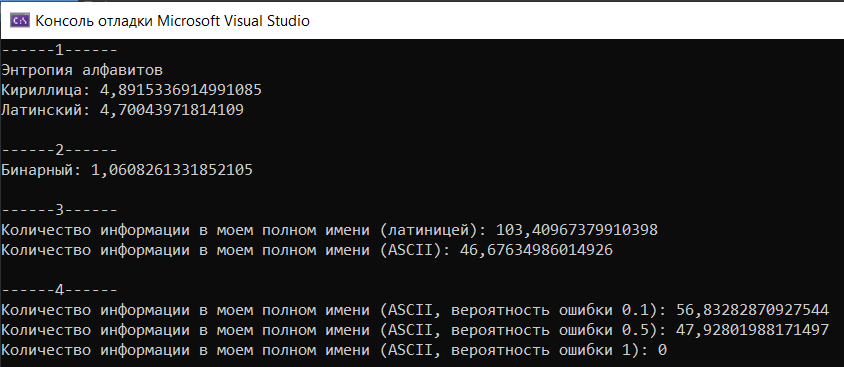


Рис. *3* – Результат выполнения программы

**2.4 Ответы на вопросы:**

1. **Что такое алфавит источника сообщения?**

**Алфавит** – конечная совокупность символов (знаков), с помощью которых можно представить любое сообщение в ИС.

1. **Что такое мощность алфавита источника сообщения?**

**Мощность алфавита** – количество символов, составляющих алфавит

1. **Какова мощность алфавита белорусского языка?**

32

1. **Какова мощность алфавита русского языка?**

33

1. **Какова мощность алфавита «компьютерного» языка?**

256

1. **Что такое энтропия алфавита?**

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является *энтропия*. С физической точки зрения энтропия показывает, какое количество информации (бит) приходится в среднем на один символ алфавита.

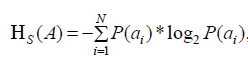
1. **От чего зависит энтропия алфавита?**

От частоты встречаемости символов и мощности алфавита.

1. **Что такое энтропия сообщения?**

Энтропия – информационная характеристика алфавита (источника сообщений на основе этого алфавита), которая показывает, какое количество информации приходится в среднем на один символ алфавита (сообщения).

1. **Записать формулу для вычисления энтропии (рисунок 1).**

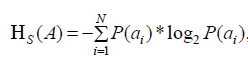


1. **Что нужно знать для вычисления энтропии алфавита?**

Частоту встречаемости каждого символа алфавита и мощность.

1. **Как рассчитываются энтропия Шеннона и энтропия Хартли? В чем принципиальное различие между этими характеристиками? Дайте толкование физического смысла энтропии.**

Энтропию алфавита А={*ai*} по К.Шеннону рассчитывают по следующей формуле, представленной на рисунке 2.



Частным случаем энтропии Шеннона является энтропия Хартли. Дополнительным условием при этом является то, что все вероятности одинаковы и постоянны для всех символов алфавита. С учетом этого формулу (2.1) можно преобразовать к виду, представленному на рисунке 3.



1. **Поясните назначение знака «минус» в формулах (2.1) и (2.4).**

Поскольку мы находим логарифмы от чисел меньших 1, то логарифм всегда будет отрицательным, поэтому мы добавляем минус перед логарифмом

1. **. Что такое избыточность алфавита и избыточность сообщений,** **сформированных в компьютерных системах? Принцип действия каких систем основан на существовании данной избыточности?**

Избыточностью алфавита называется уменьшение информационной нагрузки на один символ вследствие неравновероятности и взаимозависимости появления его символов.

Информационная избыточность характеризует относительную нагруженность алфавита.

1. **Расположите в порядке возрастания энтропии известные вам** **алфавиты.**

3,2 – белорусского

3,895 – французского

4,25 – молдавского

1. **Вычислить энтропию алфавита белорусского (русского) языка.**

3.2 бит

1. **Вычислить энтропию Шеннона бинарного алфавита, если вероятность появления в произвольном документе на основе этого алфавита одного из символов составляет 0.25, другого – 0.75; либо 0 и 1.0; либо 0.5 и 0.5.**
2. 0,811 бит
3. 0 бит
4. 1 бит
5. **Чему равна энтропия алфавита по Хартли, если мощность этого алфавита равна: а) 1 символ, б) 2 символа, в) 8 символов?**
6. 0 бит
7. 1 бит
8. 3 бит

**Вывод:** приобрела практические навыки расчета и анализа параметров и информативных характеристик дискретных ИС.

# **Лабораторная работа №3**

**Элементы теории информации. Информативность данных в различных кодировках**

**Цель:** приобретение практических навыков трансформации данных и сопоставление энтропийных свойств используемых при этом алфавитов.

**3.1 Теоретические сведения**

Из энтропийных оценок (алфавитов и сообщений), полученных в ходе выполнения лабораторной работы № 2, мы выяснили, что энтропия зависит от статических характеристик самих алфавитов и сообщений (вспомним энтропию по Шеннону и по Хартли).

Энтропия максимальна при равномерном появлении букв на любом месте сообщения. Для характеристики источника сообщений с различным алфавитом представляет интерес сравнение фактической энтропии источника с максимально возможной. В этом смысле введено понятие избыточности источника сообщений, или избыточности алфавита.

**Избыточностью алфавита** называют уменьшение информационной нагрузки на один символ вследствие разной вероятности и взаимозависимости появления его символов в сообщениях.

В наиболее общем виде избыточность алфавита R можно оценить отношением энтропии по Хартли и по Шеннону; при этом первая рассчитывается по выражению (2.2), вторая – по формуле (2.1):



При выполнении предыдущей работы мы убедились, что формально одно и то же сообщение, но представленное на основе алфавита русского (белорусского, английского или иного) языка – с одной стороны, и представленное в кодах ASCII – с другой, будут характеризоваться различным количеством содержащейся в них информации. Эта дополнительная избыточность обусловлена переносом сообщения из одной среды в другую или, иначе говоря, кодированием символов исходного алфавита.

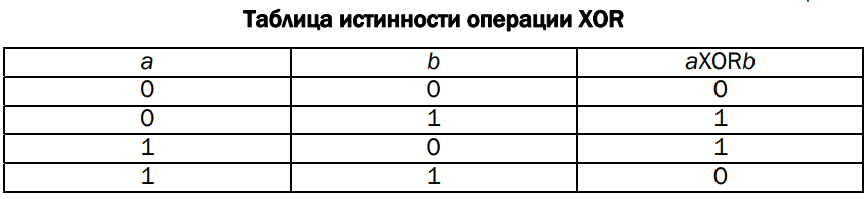
Утверждение восьмибитных кодировок (ASCII) как стандарта принесло некоторые проблемы. К этому моменту уже существовала определенная инфраструктура, использующая семибитные кодировки. Известны проблемы с «обрезанием восьмого бита» в системе электронной почты. Утверждение восьмибитного символа дало 256 различных значений, что позволило уместить в одной кодовой таблице и общепринятые символы (цифры, знаки препинания, латиницу), и символы кириллицы.

Уже созданное к тому времени и работающее программное обеспечение зачастую было приспособлено для семибитных кодировок, что приводило, например, к тому, что почтовый сервер при передаче письма обнулял старшие биты в каждом байте сообщения.

Одним из решений проблемы стала кодировка (а точнее – алгоритм) base64. В PGP алгоритм base64 используется для кодирования бинарных данных. Кодирование base64 разработано для представления произвольных последовательностей октетов в форме, позволяющей использовать строчные и прописные буквы. Используется 65-символьное подмножество набора символов US-ASCII, обеспечивающее представление одним печатным символом 6 битов данных (дополнительный 65-й символ используется для обозначения функции специальной обработки).

Kаждые 6 битов буфера, начиная с самых старших, используются как индексы строки «ABCDEFGHIJKLMNOPQRSTUV WXYZabcdefghijklmnopqrstuvwxyz0123456789+/», и ее символы, на которые указывают индексы, помещаются в выходную строку.

**Если кодируются только один или два байта, в результате получаются только первые два или три символа строки, а выходная строка дополняется двумя или одним символами «=». Это предотвращает добавление дополнительных битов к восстановленным данным.**

****

Процесс повторяется над оставшимися входными данными. Такая обработка выполняется в тех случаях, когда последняя группа входных данных содержит меньше 24 битов. Кодируемое значение всегда завершается полным квантом кодирования.

Если на входе доступно менее 24 битов, входная группа дополняется (справа) нулями до формирования целого числа 6-битных групп. Заполнение в конце данных осуществляется как раз с использованием символа «=». Поскольку входная информация base64 всегда включает целое число октетов, возможны лишь перечисленные ниже случаи:

* размер финального блока кодирования на входе кратен 24 битам, кодированный результат будет содержать целое число 4-символьных групп без заполнения символами «=»;
* размер финального блока кодирования на входе составляет 8 битов, выходной блок будет представлять 2 символа, дополненные последовательностью из двух символов заполнения «==»;
* размер финального блока кодирования на входе составляет 16 битов, выходной блок будет представлять 3 символа, дополненные символом заполнения «=».

Если а и b имеют длину более 1 бита, к примеру 1 байт, то рассматриваемая операция над ними выполняется побитово. Указанным байтам могут соответствовать символы в определенной кодировке. Положим, символу «М» (hex4d) соответствует 8-битный код 01001101 (см. табл. 3.2), а символу «а» (hex61) соответствует код 01100001, тогда операция сложения по модулю 2 этих двух бинарных кодов дает 00101100 ((hex2с), или символ «,».

**3.2 Практическое задание**

Создать собственное приложение (приветствуется!) или воспользоваться Base64-онлайн-кодировщиком, с помощью которого конвертировать произвольный документ (а) на латинице (можно использовать документ из лабораторной работы № 1) в документ (б) формата base64.

С помощью приложения, созданного в лабораторной работе № 1, получить распределение частотных свойств алфавитов по документам (а) и (б). Вычислить энтропию Хартли и Шеннона, а также избыточность алфавитов. Объяснить полученный результат.

Написать функцию, которая принимает в качестве аргументов два буфера (а и b) одинакового размера и возвращает XOR (собственная фамилия (а) и имя (b); при разной длине меньшую дополнить нулями). Входные аргументы представлять: 1) в кодах ASCII; 2) в кодах base64. Что будет результатом операции аXORbXORb?

При написании не использовать стандартные функции языка программирования. Итоговые данные сравнить с результатами использования стандартных функций языка программирования (если они есть).

**3.3 Выполнение практического задания**

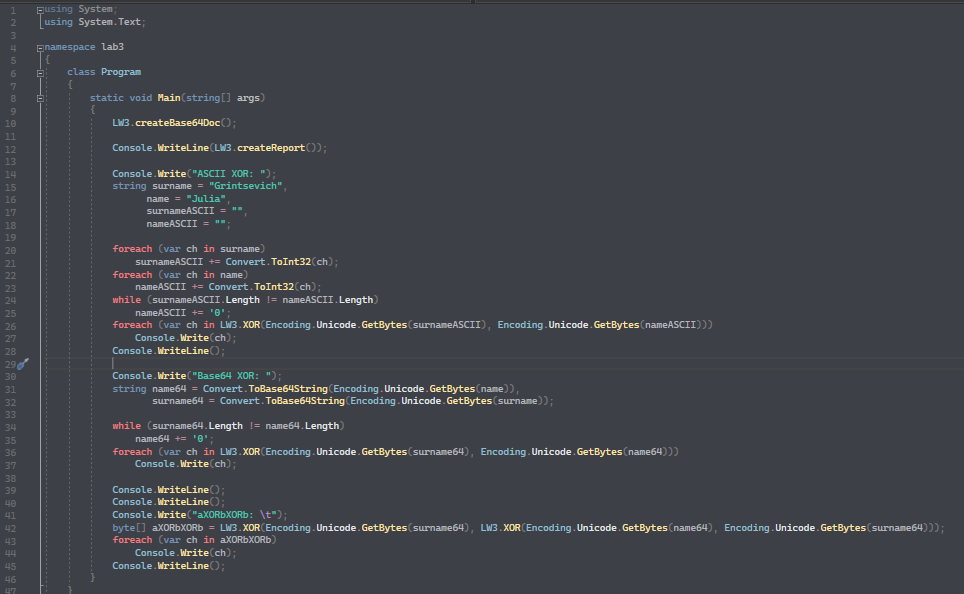
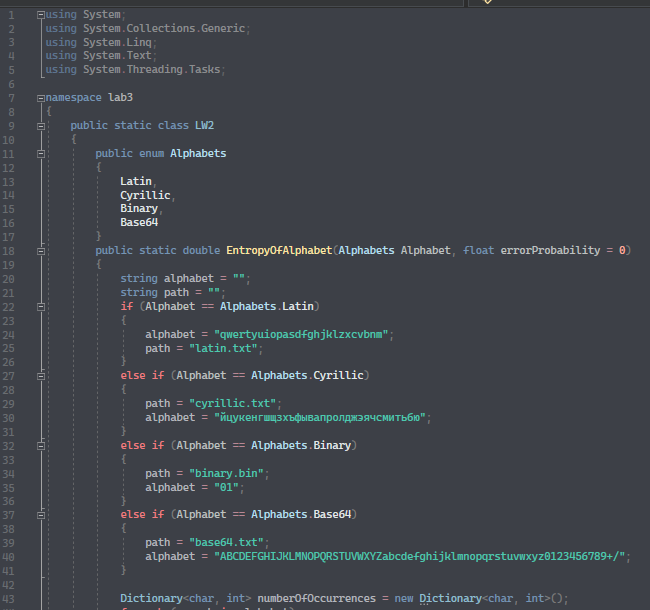
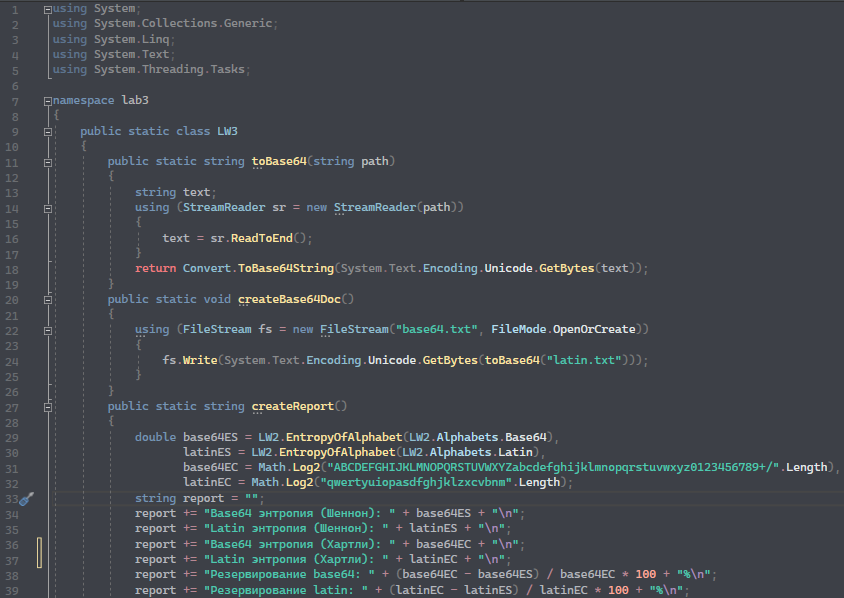
****

Рис. *1* – Код файла Program.cs

****

****

Рис. *2* – Код файла LW2.cs

****

****

Рис. *2* – Код файла LW3.cs

**3.4 Ответы на вопросы:**

1. **Что такое base64?**

**Base64** - это группа схожих binary-to-text encoding схем, которые представляют двоичные данные в ASCII-формате методом перевода в radix-64 представление. Термин Base64 происходит от a specific MIME content transfer encoding.

Кодирование Base64 широко используется в случаях, когда требуется перекодировать двоичные данные для передачи по каналу приспособленному для передачи текстовых данных.

. Используется 65-символьное подмножество набора символов US-ASCII, обеспечивающее представление одним печатным символом 6 битов данных (дополнительный 65-й символ используется для обозначения функции специальной обработки).

Избыточностью алфавита называется уменьшение информационной нагрузки на один символ вследствие неравновероятности и взаимозависимости появления его символов.

Информационная избыточность характеризует относительную нагруженность алфавита

1. **Как проверить, была ли определенная строка символов закодирована в base64?**

 Мы можем использовать следующее регулярное выражение, чтобы проверить, закодирована ли строка в base64 или нет:

**^([A-Za-z0-9+/]{4})\*([A-Za-z0-9+/]{3}=|[A-Za-z0-9+/]{2}==)?$**

В кодировке base64 набор символов - [A-Z, a-z, 0-9 и + /]. Если остальная длина меньше 4, строка заполняется символами '='.

^([A-Za-z0-9+/]{4})\* означает, что строка начинается с 0 или более групп base64.

([A-Za-z0-9+/]{4}|[A-Za-z0-9+/]{3}=|[A-Za-z0-9+/]{2}==)$ означает, что строка заканчивается одной из трех форм: [A-Za-z0-9+/]{4}, [A-Za-z0-9+/]{3}= или [A-Za-z0-9+/]{2}==.

1. **Охарактеризовать энтропийные свойства алфавитов в проанализированных форматах данных.**

*Энтропия алфавита - это количество информации, приходящееся на один символ.* Другими словами, это информационная нагрузка, которую несет один символ алфавита.

1. **Объяснить результат операции аXORbXORb. Где может найти применение такая операция?**

Эта операция называется также cложением по модулю 2, логическим сложением, исключающим «ИЛИ», строгой дизъюнкций, поразрядным дополнением.

1. **Как будут выглядеть строки: efd8b295a633908a3c0828b2 faea8766 4d72cde3aaa0 после их конвертации в base64?**
2. **Результатом операции аXORb (а – каждый байт строки, b – некоторая неизменная величина) будет строка:**

**1f180d1e1f04051c404c0f19**

**1f180308050d024c030a4c18**

**04094c1f18030009024c1c00**

**Найти значение b.**

# **Лабораторная работа №4**

**Избыточное кодирование данных в информационных системах. Код Хемминга**

**Цель работы**: приобретение практических навыков кодирования/декодирования двоичных данных при использовании кода Хемминга.

**4.1 Теоретические сведения**

**Надежность системы** – характеристика способности программного, аппаратного, аппаратно-программного средства выполнить при определенных условиях требуемые функции в течение конкретного периода времени.

**Достоверность работы системы (устройства**) – свойство, характеризующее истинность конечного (выходного) результата работы (выполнения программы), определяемое способностью средств контроля фиксировать правильность или ошибочность работы.

**Избыточное кодирование —** вид кодирования, использующий избыточное количество информации с целью последующего контроля целостности данных при записи/воспроизведении информации или при её передаче по линиям связи.

**Код Хемминга** - это алгоритм, который позволяет закодировать какое-либо информационное сообщение определённым образом и после передачи (например по сети) определить появилась ли какая-то ошибка в этом сообщении (к примеру из-за помех) и, при возможности, восстановить это сообщение.

**Ошибка устройства** – неправильное значение сигнала (бита – в цифровом устройстве) на внешних выходах устройства или отдельного его узла, вызванное технической неисправностью, или воздействующими на него помехами (преднамеренными либо непреднамеренными), или иным способом.

**Ошибка программы** – проявляется в не соответствующем реальному (требуемому) промежуточном или конечном значении(результате) вследствие неправильно запрограммированного алгоритма или неправильно составленной программы.

*Надежность является* комплексным свойством, включающим в себя единичные свойства: безотказность, ремонтопригодность, сохраняемость, долговечность.

**Безотказность** – это свойство технического объекта непрерывно сохранять работоспособное состояние в течение некоторого времени (или наработки). Наработка, как правило, измеряется в единицах времени.

**Ремонтопригодность** – это свойство технического объекта, заключающееся в приспособленности к поддержанию и восстановлению работоспособного состояния путем технического обслуживания, ремонта (или с помощью дополнительных, избыточных технических средств, функционирующих параллельно с объектом).

Код Хемминга относится к классу линейных блочных кодов.

**Линейные блочные коды** – это класс кодов с контролем четности, которые можно описать парой чисел (п, k).

Для формирования r проверочных символов (кодирования), т. е. вычисления проверочного слова Xr, используется *порождающая матрица G:* совокупность базисных векторов будем далее записывать в виде матрицы G размерностью k×n с единичной подматрицей (I) в первых k строках и столбцах:



Более *точно матрица G называется* порождающей матрицей линейного корректирующего кода в приведенно-ступенчатой форме. Кодовые слова являются линейными комбинациями строк матрицы G (кроме слова, состоящего из нулевых символов).

Кодирование заключается в умножении вектора сообщения Хk длиной k на порождающую матрицу по правилам матричного умножения (все операции выполняются по модулю 2). Очевидно, что при этом первые k символов кодового слова равны соответствующим символам сообщения, а последние r символов образуются как линейные комбинации первых [5, 9].

Для всякой порождающей матрицы G существует матрица Н размерности r×n, задающая базис нулевого пространства кода и удовлетворяющая равенству



Справедливо также



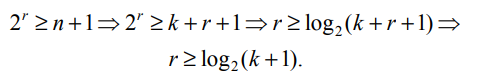
В последнем выражении символ «T» означает транспонирование, а X n = 1x , 2 x , …, xn. Матрица Н, называемая проверочной, равна



Количество r избыточных (проверочных) символов кодового слова определяется из следующей простой логической цепи рассуждений. Общее число всех возможных комбинаций 2r должно удовлетворять неравенству



в силу того, что



Присутствие цифры «1» в приведенных выражениях соотносит ее с нулевым вектор-столбцом, который в матрице не используется.

Результат умножения сообщения на выходе канала передачи (Yn) или (что равнозначно) сообщения, считываемого из памяти, на проверочную матрицу (Н) **называется синдромом (вектором ошибки) S:**



где Yn = y1, y2, …, yn – принятый вектор (сообщение на выходе канала), полученный после передачи либо считывания из памяти. Вектор Yn обычно представляют в следующем виде:



где Еn = е1, е2, …, еn – вектор ошибки.

**Синдром** – это результат проверки четности, выполняемой над сообщением Yn для определения его принадлежности заданному набору кодовых слов. При положительном результате проверки синдром S равен 0, т. е. Yn = Хn. Если Yn содержит ошибки, которые можно исправить, то синдром имеет определенное ненулевое значение, что позволяет обнаружить и исправить конкретную ошибочную комбинацию.

Важно запомнить, что в силу выражений (4.3)–(4.7) ненулевой синдром всегда равен сумме по модулю 2 тех векторстолбцов матрицы Н, номера которых соответствуют номерам ошибочных битов в слове Yn.

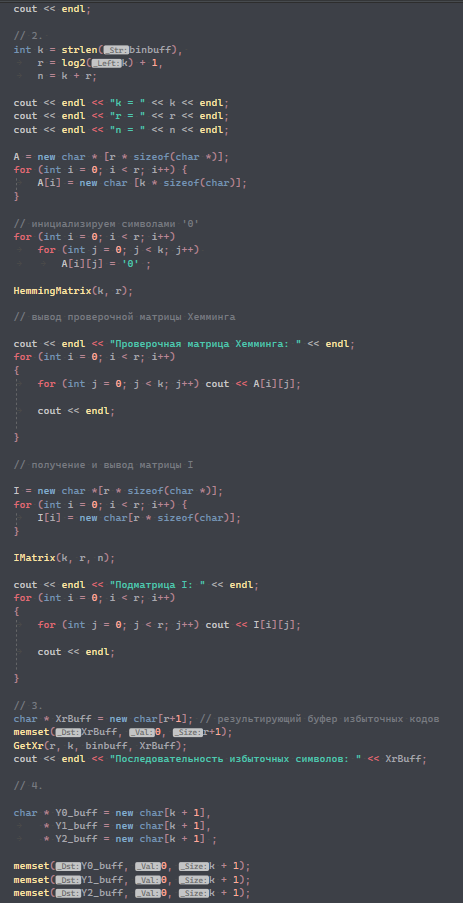
При этом нужно помнить, что вид матрицы (4.8) не соответствует ее каноническому представлению, поскольку во всех столбцах единичной матрицы, кроме последнего, будет по 2 единицы.

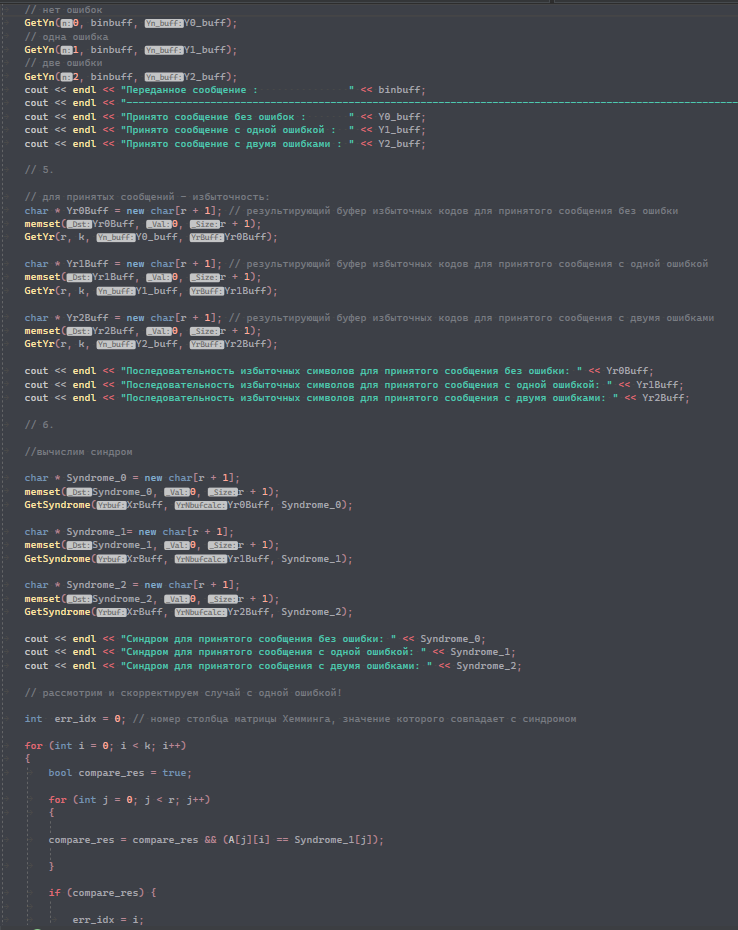
**4.2 Практическое задание**

1. На основе информационного сообщения, представленного символами русского/английского алфавитов, служебными символами и цифрами, содержащегося в некотором текстовом файле (согласовать с преподавателем), сформировать информационное сообщение в двоичном виде; длина сообщения в бинарном виде должна быть не менее 16 символов. Для выполнения этого задания можно использовать коды ASCII символов алфавита либо результаты лабораторной работы № 3.
2. Для полученного информационного слова построить проверочную матрицу Хемминга (значение минимального кодового расстояния согласовать с преподавателем).
3. Используя построенную матрицу, вычислить избыточные символы (слово Xr).
4. Принять исходное слово со следующим числом ошибок: 0, 1, 2. Позиция ошибки определяется (генерируется) случайным образом.
5. Для полученного слова Yn = Yk, Yr, используя уже известную проверочную матрицу Хемминга, вновь вычислить избыточные символы (обозначим их Yr’), используя выражение (4.6).
6. Вычислить и проанализировать синдром. В случае, если анализ синдрома показал, что информационное сообщение было передано с ошибкой (или 2 ошибками), сгенерировать унарный вектор ошибки Еn = е1, е2, …, еn и исправить одиночную ошибку, используя формулу (4.7); проанализировать ситуацию при возникновении ошибки в 2 битах.
7. Результаты оформить в виде отчета по установленным правилам. Файл анализируемой информации (а соответственно и интерфейс приложения) должен содержать исходное информационное сообщение, значения величин k, r, n, проверочную матрицу Хемминга Hn,k, слово Xn, Xr, Yn,Yr, Yr’, синдром S, вектор ошибки Еn. Программа не должна быть чувствительна к длине информационного сообщения.

**4.3 Выполнение практического задания**

****

****

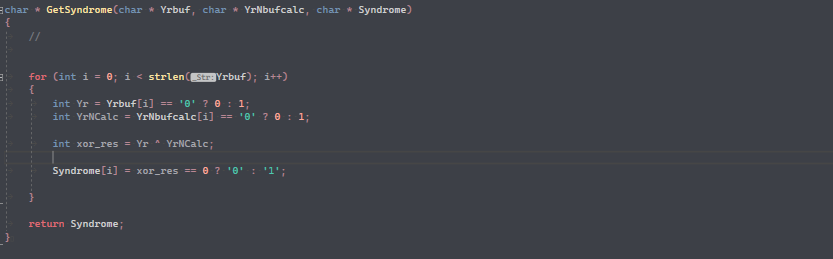
****

****

****

****

****

****

****

**4.4 Ответы на вопросы:**

1. *В чем заключается цель и функциональная сущность преобразования информации на основе избыточного кодирования?*

Сущность методов избыт. кодирования инф. основывается на том, что произвольное сообщение Хк, состоящее из К символов (так наз. информационные символы), дополняется словом Хr, состоящ. из R символов (проверочные, =контрольные символы). Они избыточные.

Вместо К симв. передается N=(К+R) симв. Общее слово Хn называется кодовым словом. R=f(К, корректирующ.код).

1. *Пояснить зависимость r от длины информационного слова k. Охарактеризовать относительную избыточность сообщения и время его передачи по сети.*

r = n – k

1. *Предположим, есть выбор (при построении матрицы кода) между вектор-столбцами большего и меньшего веса. Какой вариант Вы предпочтете и почему?*

Меньшего.

# **Лабораторная работа №5**

**Избыточное кодирование данных в информационных системах. Итеративные коды**

**Цель работы**: приобретение практических навыков кодирования/декодирования двоичных данных при использовании итеративных кодов.

**Задачи:**

1. Закрепить теоретические знания по использованию итеративных кодов для повышения надежности передачи и хранения в памяти компьютера двоичных данных.

2. Разработать приложение для кодирования/декодирования двоичной информации итеративным кодом с различной относительной избыточностью кодовых слов.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**5.1. Теоретические сведения**

Избыточное кодирование используется для увеличения надежности.

Итеративные коды относятся к классу кодов произведения.

**Кодом произведения двух исходных (базовых) помехоустойчивых кодов** называется такой многомерный помехоустойчивый код, кодовыми последовательностями которого являются все двумерные таблицы со строками кода (k1) и столбцами кода (k2).

Паритет – избыточный символ. Записываются сверху вниз, справа налево. Возможен обратный или иной порядок. Важно только, чтобы при декодировании сообщения использовался аналогичный порядок следования паритетов. Символ Xhv (**паритет паритетов**) равен сумме по модулю 2 символов информационного слова Xk, а также проверочных символов Xv и Xh.

Паритет паритетов - равен сумме по модулю 2 символов информационного слова Xk, а также проверочных символов Xv и Xh.

Итеративные коды могут строиться на основе использования дву-, трехмерных матриц (таблиц) и более высоких размерностей. Каждая из отдельных последовательностей информационных символов кодируется определенным линейным кодом (групповым или циклическим). Получаемый таким образом итеративный код также является **линейным**.

Простейшим из итеративных кодов является **двумерный** код с проверкой на четность по строкам и столбцам. **Итеративные коды**, иногда называемые прямоугольными кодами (англ. rectangular code) либо композиционными (англ. product code), являются одними из самых простых (с точки зрения аппаратной реализации) избыточных кодов, позволяющих исправлять ошибки в информационных словах.

**Основное достоинство рассматриваемых кодов – простота как аппаратной, так и программной реализации. Основной недостаток – сравнительно высокая избыточность.**

В упомянутой двумерной матрице кодовые слова записываются в виде таблицы. Проверочные символы вычисляются исходя из того, что строки и столбцы должны содержать четное (нечетное) число единиц. Например, при кодировании информационного слова Хk = 011101111 с помощью таблицы с четностью по строкам и столбцам получим избыточные символы Хr = Xh, Xv, Xhv = 0010011, как показано на рис. 5.1 (информационные символы выделены жирным шрифтом, а проверочные – курсивом).

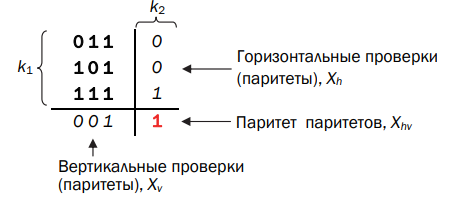


Рис. *5*.*1* – Пояснение к принципу формирования избыточных символов итеративного кода

В соответствии с рис.1 кодовое слово будет иметь следующий вид: Xn = **0111011110010011**. Как видно, избыточные символы (называемые также паритетами) в приведенном кодовом слове в принятом порядке (Xh, Xv, Xhv) записываются сверху вниз, справа налево. Возможен обратный или иной порядок. Важно только, чтобы при декодировании сообщения использовался аналогичный порядок следования паритетов. Символ Xhv (паритет паритетов) равен сумме по модулю 2 символов информационного слова Xk, а также проверочных символов Xv и Xh.

Поскольку двумерная матрица формируется как комбинация двух кодов простой четности (по каждому измерению), каждый из которых характеризуется минимальным кодовым расстоянием dmin = 2, то полученный итеративный код (r = k1 + k2) будет характеризоваться минимальным кодовым расстоянием, равным произведению dmin по строкам и по столбцам, т. е. 4.

Использование символа Xhv обеспечивает минимальное кодовое расстояние такого итеративного кода dmin (r = k1 + k2 + 1) на единицу больше. В этом легко обнаруживается сходство кода с кодом Хемминга при dmin = 4.

Нетрудно также представить процесс вычисления проверочных символов кодового слова для примера на рис. 5.1 с помощью проверочной матрицы Хемминга и соотношения (4.4). Для указанного примера проверочная матрица кода с dmin = 3 выглядит так:

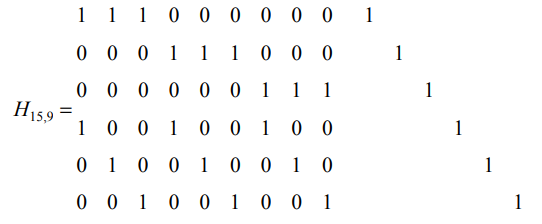


Рис. 5.2 – Проверочная матрица

Передачу символов кода обычно осуществляют последовательно символ за символом, от одной строки к другой, либо параллельно целыми строками. Как показано на рис. 5.1, проверочный символ есть сверка по модулю 2 информационных символов, записанных в соответствующие строку или столбец матрицы. Декодирование начинают сразу, не ожидая поступления всего блока информации. Проверка соответствия избыточных символов полученного слова (Yr = Yh, Yv, Yhv либо Yr = Yh, Yv) при декодировании позволяет обнаружить любое нечетное число искаженных символов, расположенных в одной строке или в одном столбце. Формально такое декодирование осуществляется сравнением принятых (Yh, Yv, Yhv) и вновь вычисленных (Y’h, Y’v, Y’hv) для полученного слова паритетов. В упрощенной форме это показано на рис. 3. Определение местоположения одиночной ошибки по строке указывает на наличие ошибки в этой строке матрицы, а проверка по столбцу – конкретный символ (рис. 3, а).

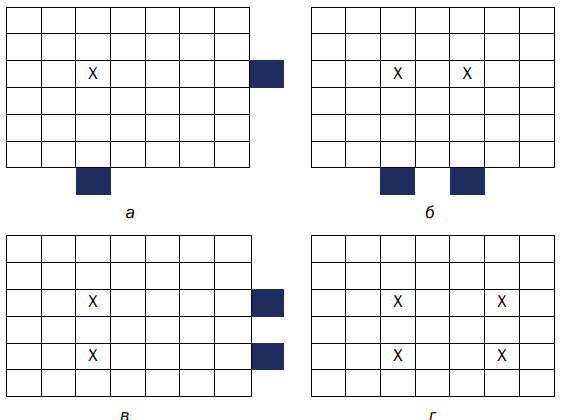


Рис. *5*.*3* – Пояснение к принципу обнаружения местоположения ошибочных битов в принятом сообщении

Однако этим кодом не могут быть установлены местоположения многократных ошибок, имеющих четное число искаженных символов как по строкам, так и по столбцам (рис. 3, б, в). Простейшая необнаруживаемая ошибка содержит четыре искаженных символа, расположенных в вершинах прямоугольника или квадрата (рис. 3, г). Это происходит из-за того, что четность (паритет) по строкам и по столбцам матрицы не нарушается. Полезную информацию о кодировании и декодировании информации итеративным кодом можно найти в источниках.

**5.2 Практическое задание**

Разработать собственное приложение, которое позволяет выполнять следующие операции:

1) вписывать произвольное двоичное представление информационного слова Хk (кодируемой информации) длиной k битов в двумерную матрицу размерностью в соответствии с вариантом либо в трехмерную матрицу в соответствии с вариантом (указаны в табл. 5.1);

2) вычислять проверочные биты (биты паритетов): а) по двум; б) по трем; в) по четырем направлениям (группам паритетов);

3) формировать кодовое слово Xn присоединением избыточных символов к информационному слову;

4) генерировать ошибку произвольной кратности (i, i > 0), распределенную случайным образом среди символов слова Xn, в результате чего формируется кодовое слово Yn;

5) определять местоположение ошибочных символов итеративным кодом в слове Yn в соответствии с используемыми группами паритетов по пункту (2) и исправлять ошибочные символы (результат исправления – слово Yn’);

6) выполнять анализ корректирующей способности используемого кода (количественная оценка) путем сравнения соответствующих слов Xn и Yn’; результат анализа может быть представлен в виде отношения общего числа сгенерированных кодовых слов с ошибками определенной одинаковой кратности (с одной ошибкой, с двумя ошибками и т. д.) к числу кодовых слов, содержащих ошибки этой кратности, которые правильно обнаружены и которые правильно скорректированы.

Таблица 5.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Вариант | Длина информационного слова (бит), *k* | *k*1 | *k*2 | *z* | Количество групп паритетов |
| 1 | 16 | 4 8 4 2 | 4 2 2 4 | - – 2 2 | 2; 3 2; 3 2; 3; 4; 5 2; 3; 4; 5 |
| 2 | 20 | 4 2 2 2 | 5 10 5 2 | – – 2 5 | 2; 3 2; 3 2; 3; 4; 5 2; 3; 4; 5 |
| 3 | 24 | 4 3 3 6 | 6 8 3 2 | – – 4 2 | 2; 3 2; 3 2; 3; 4; 5 2; 3; 4; 5 |
| 4 | 32 | 4 2 8 4 | 8 16 2 4 | – – 2 2 | 2; 3 2; 3 2; 3; 4; 5 2; 3; 4; 5 |
| 5 | 40 | 5 4 5 2 | 8 10 4 10 | – – 2 2 | 2; 3 2; 3 2; 3; 4; 5 2; 3; 4; 5 |

**5.3 Выполнение практического задания**



Рис. *5*.*4* – Код файла Program.cs



Рис. *5*.5 – Код файла Program.cs



Рис. *5*.*6* – Код файла Program.cs

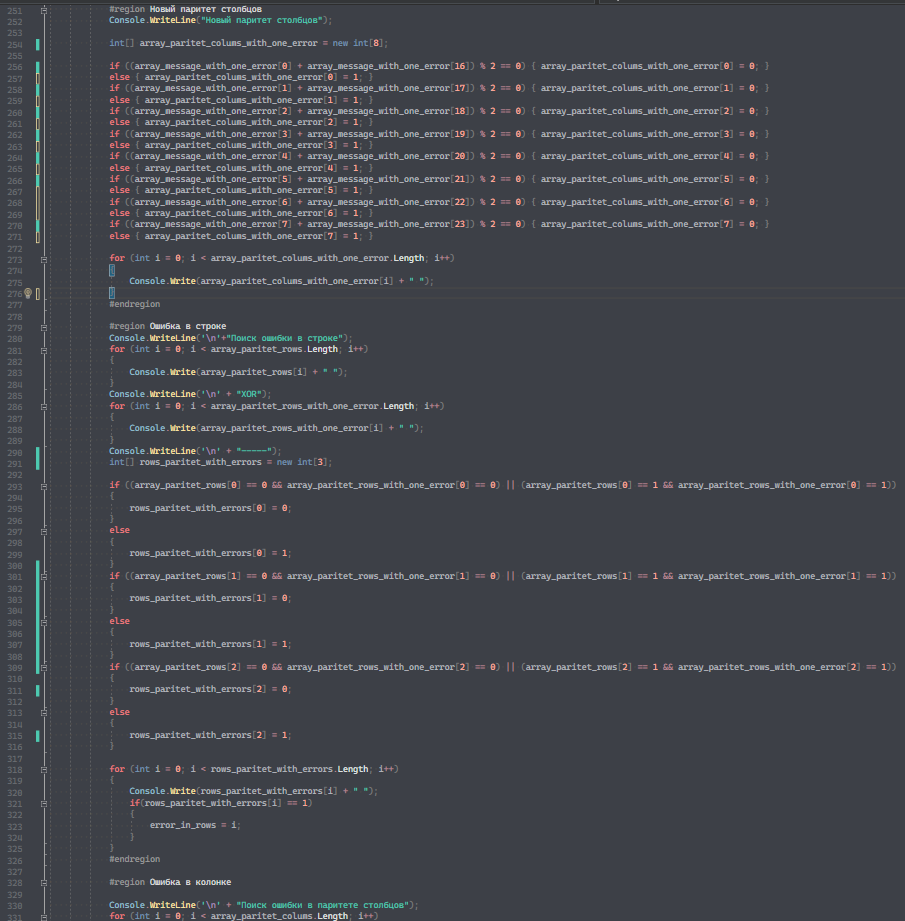
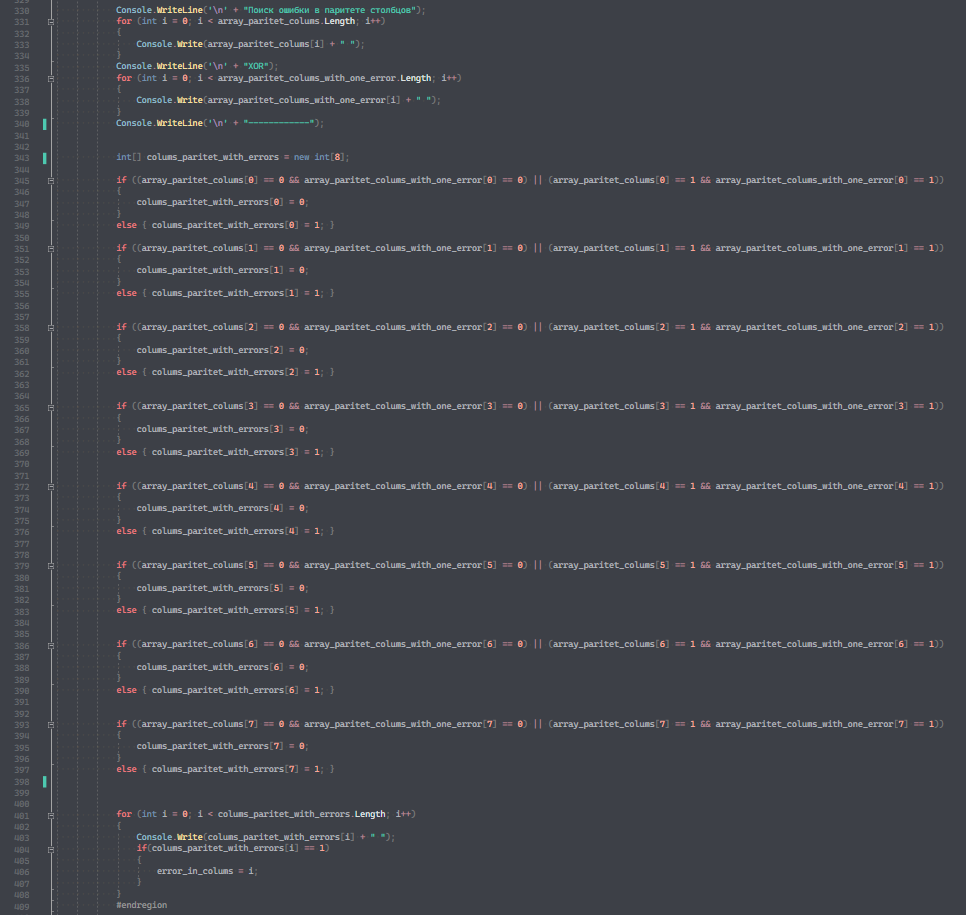


Рис. *5*.7 – Код файла Program.cs



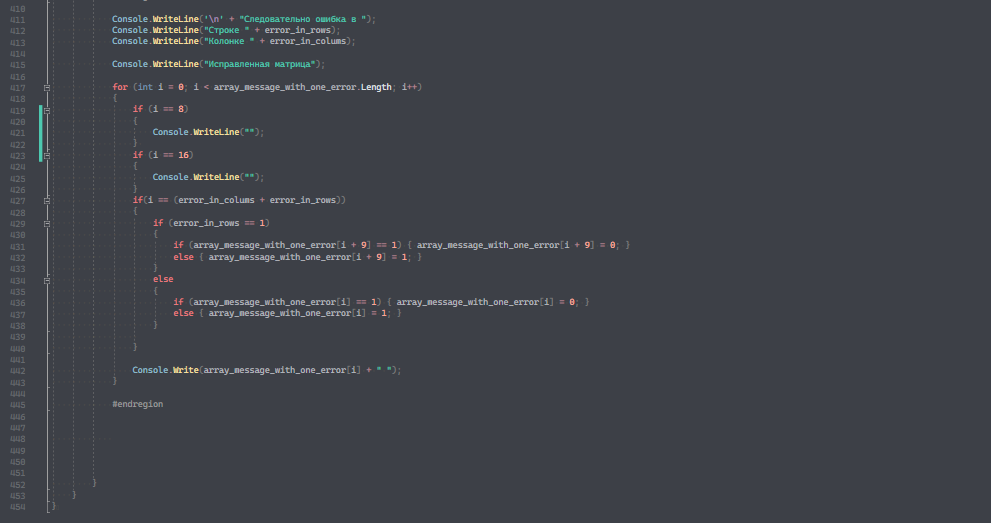


Рис. *5*.*8* – Код файла Program.cs

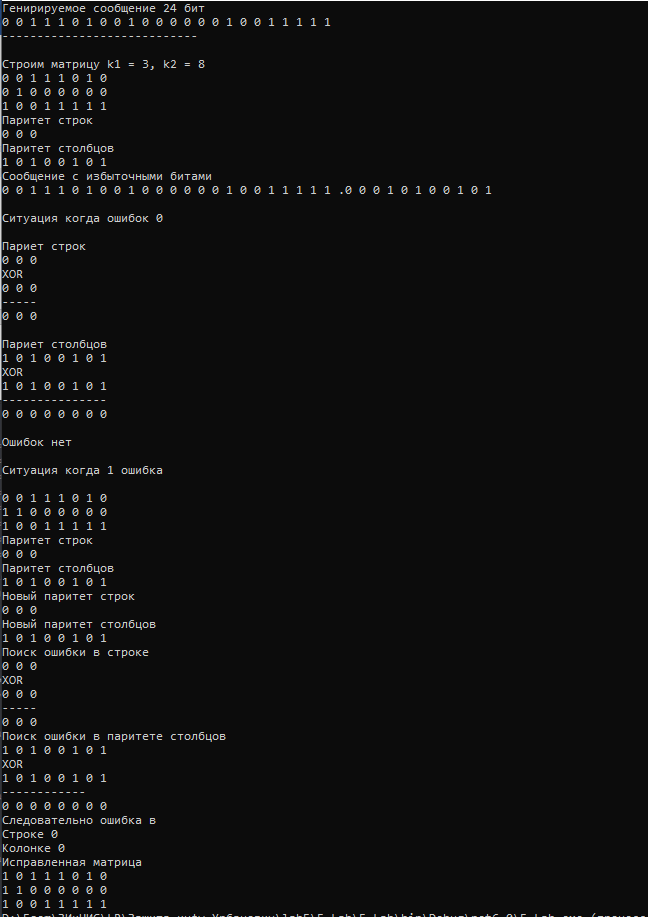


Рис. *5*.*9* – Результат выполнения программы

**5.4 Ответы на вопросы:**

1. *Охарактеризовать основные параметры итеративного кода.*

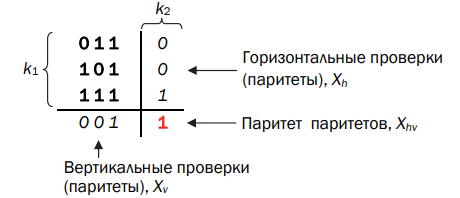
Итеративные помехоустойчивые коды относятся к классу кодов произведения. Кодом произведения двух исходных (базовых) помехоустойчивых кодов n1и n2называется такой многомерный помехоустойчивый код N=n1\*n2, кодовыми последовательностями которого являются все двумерные таблицы со строками кода n1и столбцами кода п2.

Итеративные коды могут строиться на основе использова­ния двух, трехмерных матриц (таблиц) и более высоких размерностей.

Простейшим из итеративных кодов является двумерный код с проверкой на четность по строкам и столбцам.

Основное достоинство рассматриваемых кодов – простота как аппаратной, так и программной реализации. Основной недостаток – сравнительно высокая избыточность.

Проверочные символы вычисляются исходя из того, что строки и столбцы должны содержать четное (нечетное) число единиц.



Паритет – избыточный символ. Записываются сверху вниз, справа налево. Возможен обратный или иной порядок. Важно только, чтобы при декодировании сообщения использовался аналогичный порядок следования паритетов. Символ Xhv (паритет паритетов) равен сумме по модулю 2 символов информационного слова Xk, а также проверочных символов Xv и Xh.

1. *Какое максимальное число ошибок может быть обнаружено итеративным кодом? При каком условии?*

Все ошибки кратности 4 и менее.

1. *Определить, какая геометрическая фигура, являющаяся формой для записи символов информационного слова, обеспечивает наименьшую относительную избыточность кодового слова при фиксированном (каком?)*

k.

# **Лабораторная работа №6**

**Избыточное кодирование данных в информационных системах. Циклические коды**

**Цель работы**: приобретение практических навыков кодирования/декодирования двоичных данных при использовании циклических кодов.

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию ЦК для повышения надежности передачи и хранения в памяти компьютера двоичных данных, для контроля интегральности файлов информации.

2. Разработать приложение для кодирования/декодирования двоичной информации циклическим кодом.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**6.1. Теоретические сведения**

**Циклические коды** − это семейство помехоустойчивых кодов, линейный, блочный код, одной из разновидностей которых являются коды Хемминга.

**Основные свойства ЦК:**

относятся к классу линейных, систематических;

сумма по модулю 2 двух разрешенных кодовых комбинаций дает также разрешенную кодовую комбинацию;

каждый вектор (кодовое слово), получаемый из исходного кодового вектора путем циклической перестановки его символов, также является разрешенным кодовым вектором; к примеру, если кодовое слово имеет следующий вид: 1101100, то разрешенной кодовой комбинацией будет и такая: 0110110;

при простейшей циклической перестановке символы кодового слова перемещаются слева направо на одну позицию, как в приведенном примере;

поскольку к числу разрешенных кодовых комбинаций ЦК относится нулевая комбинация 000...00, то минимальное кодовое расстояние dmin для ЦК определяется минимальным весом разрешенной кодовой комбинации;

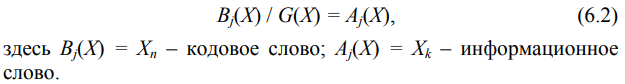
циклический код не обнаруживает только такие искаженные помехами кодовые комбинации, которые приводят к появлению на стороне приема других разрешенных комбинаций этого кода;

в основе описания и использования ЦК лежит полином или многочлен некоторой переменной (обычно Х).

Рассматриваемые операции сводятся к известным процедурам умножения и деления двоичных чисел либо соответствующих этим числам полиномов. Действия с кодовыми словами производятся по правилам арифметики по модулю 2. Следует помнить, что вычитание равносильно сложению.

**Порождающие полиномы циклических кодов**. Характеризуя ЦК в общем случае, обычно отмечают следующее: ЦК составляют множество многочленов {Вj(X)} степени r (r − число проверочных символов в кодовом слове), кратных порождающему (образующему) полиному G(Х) степени r, который должен быть делителем бинома Xn + 1, т. е. остаток после деления бинома на G(X) должен быть нулевым.

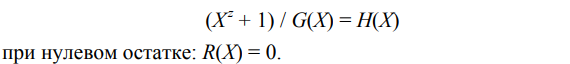
Формирование разрешенных кодовых комбинаций ЦК Bj(X) основано на предварительном выборе **порождающего (генераторного или образующего) полинома G(X)**, который обладает важным отличительным признаком: **все комбинации Bj(X) делятся на порождающий полином G(X) без остатка:**



**Степень порождающего полинома определяет число проверочных символов**: r = n – k. Из этого свойства следует простой способ формирования разрешeнных кодовых слов ЦК − умножение информационного слова A(X) на порождающий полином G(X):



**Порождающими могут быть только такие полиномы, которые являются делителями двучлена (бинома) Хz + 1:**

****

****

**Синдромом ошибки** в этих кодах **является** наличие остатка от деления принятой кодовой комбинации на порождающий полином. Если синдром равен нулю, то считается, что ошибок нет. В противном случае с помощью полученного синдрома можно определить номер разряда принятой кодовой комбинации, в котором произошла ошибка, и исправить ее примерно по той же схеме, которую мы использовали для кода Хемминга.

**Кодирование информационного слова**. Деление полиномов позволяет представить кодовые слова в виде блочного кода, т. е. информационных Хk (Аi(Х)) и проверочных Хr (Ri(X)) символов. Поскольку число последних равно r, то для компактной их записи в младшие разряды кодового слова надо предварительно к кодируемому (информационному) слову Аi(Х) справа дописать r нулевых символов.

**Декодирование принятого сообщения по синдрому. Основная операция:** принятое кодовое слово (Yn) нужно поделить на порождающий полином, который использовался при кодировании. Если Yn принадлежит коду, т. е. слово не искажено помехами, то остаток от деления (синдром) будет нулевым. Ненулевой остаток свидетельствует о наличии ошибок в принятой кодовой комбинации: Yn ≠ Хn. Для исправления ошибки нужно определить вектор (полином) ошибки Еn.

**Здесь, как и в предыдущих лабораторных работах, всякому ненулевому синдрому соответствует определенное расположение (конфигурация) ошибок: синдром для ЦК имеет те же свойства, что и для кода Хемминга.**

**Декодирование синдрома и исправление ошибки в принятом сообщении**. Декодирование ненулевого синдрома имеет целью определение ошибочного бита в принятом сообщении или, иначе говоря, определение вектора Еn. Как и в лабораторной работе № 4, поиск ошибочного бита будем производить через поиск соответствия между синдромом и проверочной матрицей кода. Наряду с полиномиальным способом задания кода, структуру построения кода можно определить с помощью матричного представления. При этом в ряде случаев проще реализуется построение кодирующих и декодирующих устройств ЦК.

**Вспомним, что в силу выражений (4.3)–(4.7) ненулевой синдром всегда равен сумме по модулю 2 тех вектор столбцов матрицы Н, номера которых соответствуют номерам ошибочных битов в слове Yn.**

**6.2 Практическое задание**

1. Задание выполняется по указанию преподавателя в соответствии с вариантом из табл. 6.2, из которого выбирается порождающий полином ЦК, а по значению соответствующего ему значения r – длина k информационного слова Xk. Полагаем, что каждый полином соответствует коду, обнаруживающему и исправляющему одиночные ошибки в кодовых словах. Определить параметры (n, k)-кода для своего варианта. Основой задания является разработка приложения.

2. Составить порождающую матрицу (n, k)-кода в соответствии с формулой (6.7), трансформировать ее в каноническую форму и далее – в проверочную матрицу канонической формы.

3. Используя порождающую матрицу ЦК, вычислить избыточные символы (слово Xr) кодового слова Xn и сформировать это кодовое слово.



4. Принять кодовое слово Yn со следующим числом ошибок: 0; 1; 2. Позиция ошибки определяется (генерируется) случайным образом.

5. Для полученного слова Yn вычислить и проанализировать синдром. В случае, если анализ синдрома показал, что информационное сообщение было передано с ошибкой (или 2 ошибками), сгенерировать унарный вектор ошибки Еn = е1, е2, …, еn и исправить одиночную ошибку, используя выражение (6.5); проанализировать ситуацию при возникновении ошибки в 2 битах.

**6.3 Выполнение практического задания**

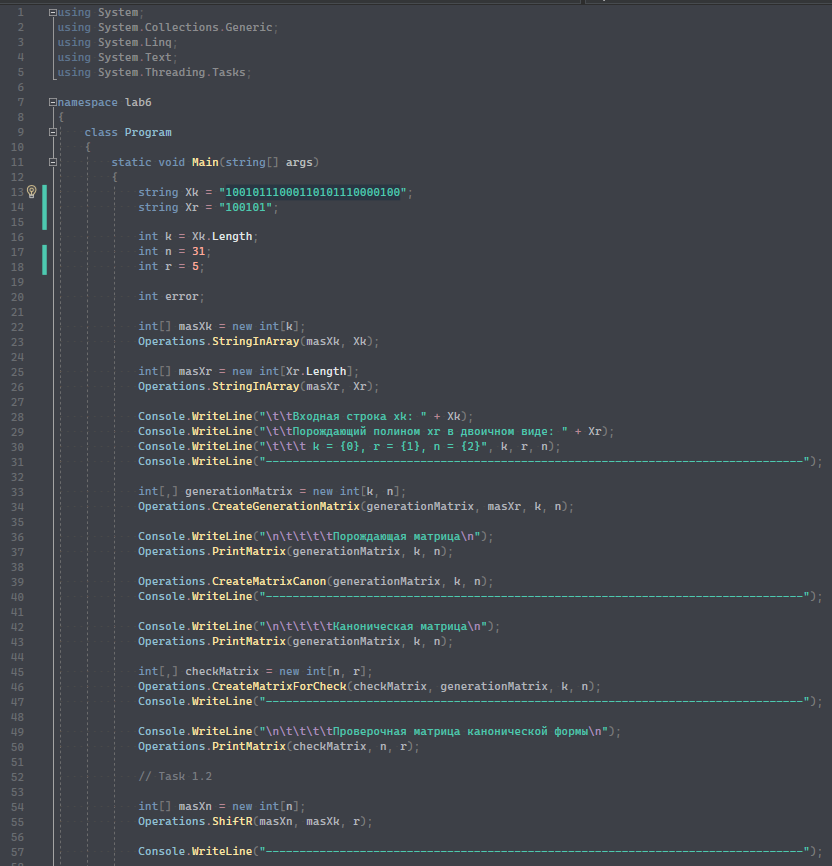


Рис. *6*.*1* – Код файла Program.cs

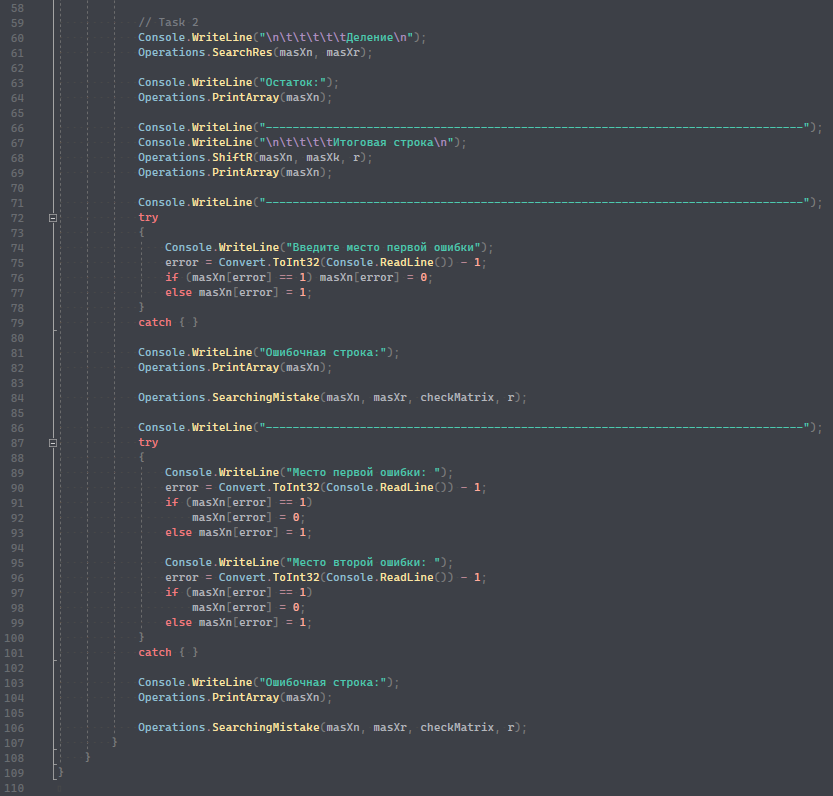


Рис. *6*.2 – Код файла Program.cs



Рис. *6*.*3* – Код файла Operations.cs

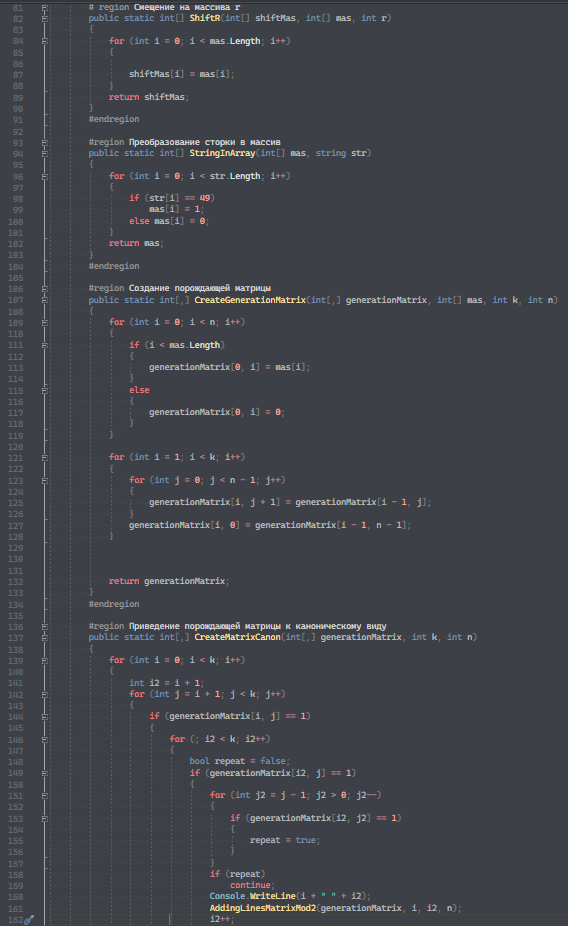


Рис. *6*.*4* – Код файла Operations.cs



Рис. *6*.*5* – Код файла Operations.cs

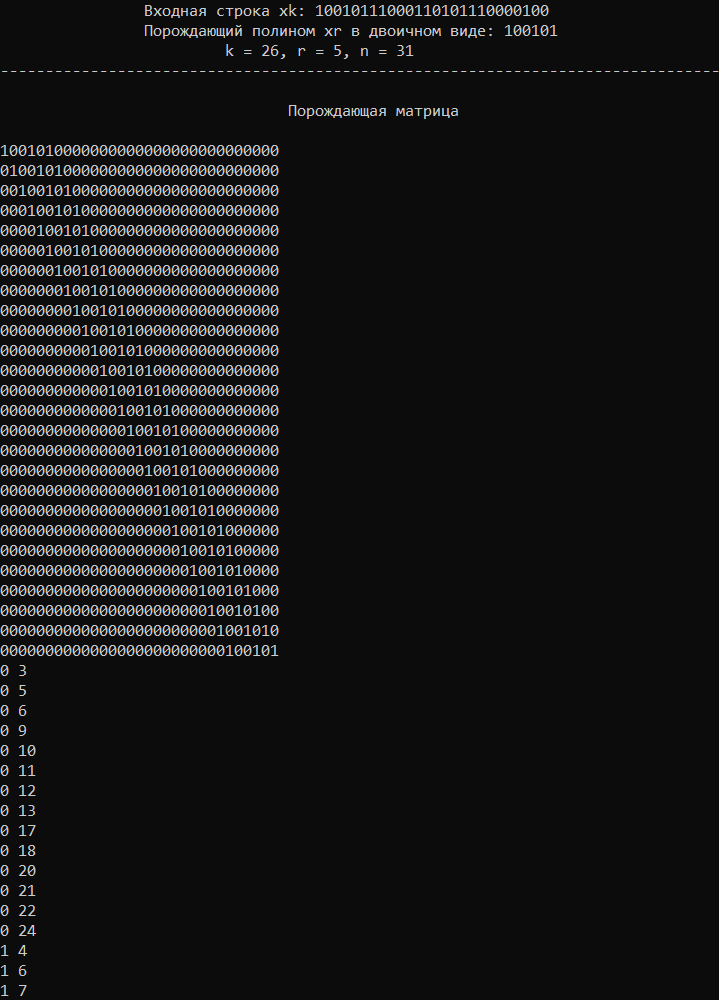
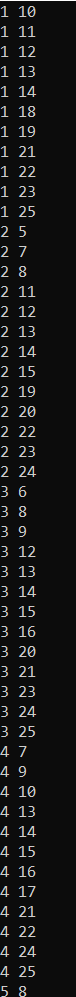
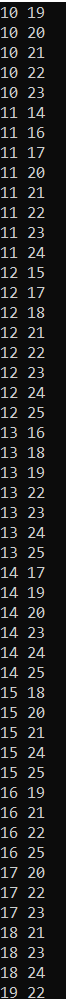
   

Рис. *6.6* – Результат выполнения программы

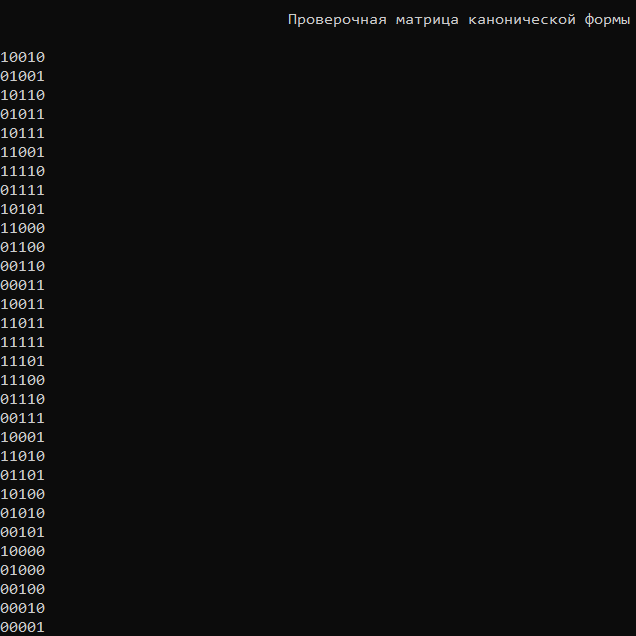
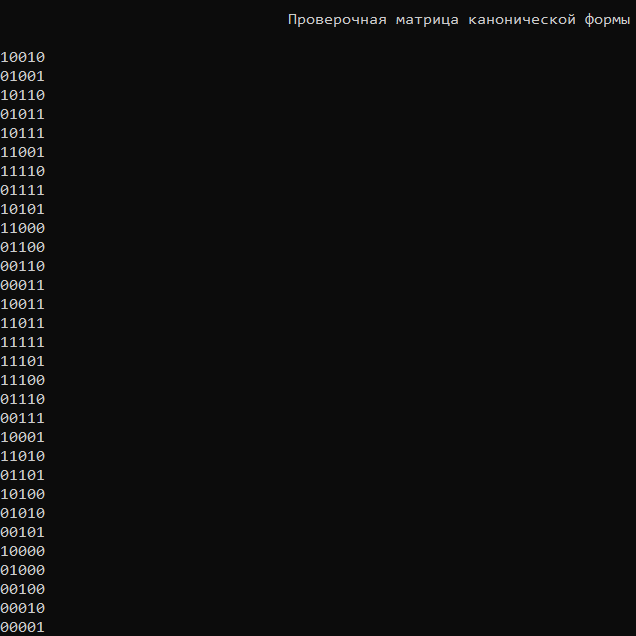
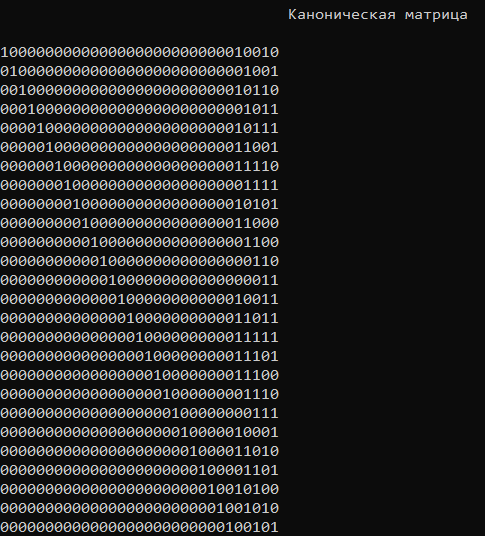


Рис. *5*.*9* – Результат выполнения программы

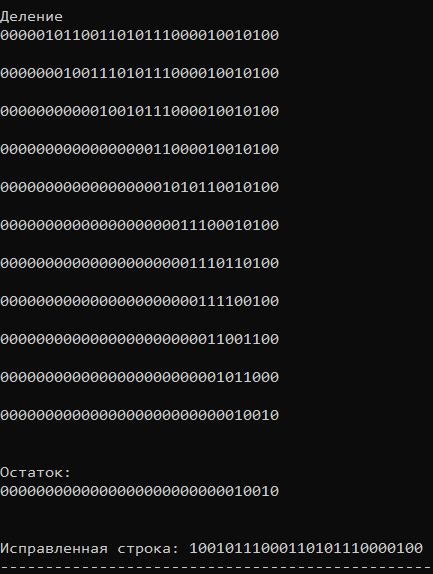
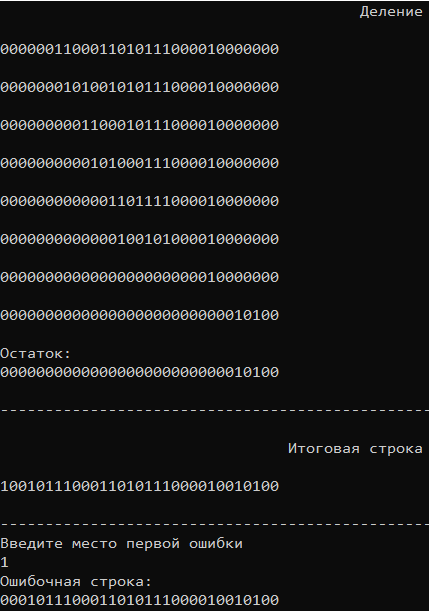


Рис. *6.7* – Результат выполнения программы

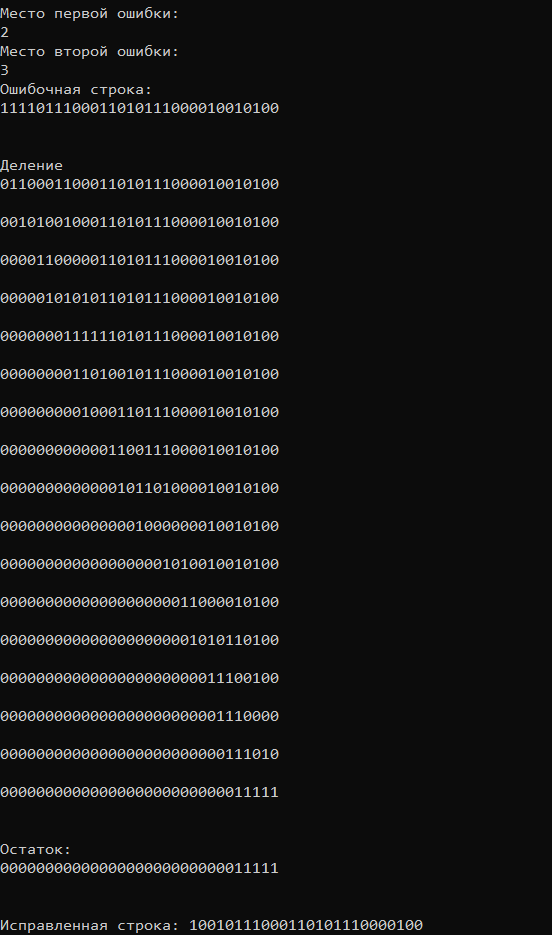


Рис. *6.9* – Результат выполнения программы

**6.4 Ответы на вопросы**

1. *Охарактеризовать основные параметры циклических кодов.*

Циклические коды − это семейство помехоустойчивых кодов, одной из разновидностей которых являются коды Хемминга.

Основные свойства ЦК:

относятся к классу линейных, систематических;

сумма по модулю 2 двух разрешенных кодовых комбинаций дает также разрешенную кодовую комбинацию;

каждый вектор (кодовое слово), получаемый из исходного кодового вектора путем циклической перестановки его символов, также является разрешенным кодовым вектором; к примеру, если кодовое слово имеет следующий вид: 1101100, то разрешенной кодовой комбинацией будет и такая: 0110110;

при простейшей циклической перестановке символы кодового слова перемещаются слева направо на одну позицию, как в приведенном примере;

1. *Пары десятичных чисел: 14, 11; 19, 15; 29, 13; 35, 45 преобразовать в двоичные числа и представить их в виде полиномов.*

14 = 1110 = x3 + x2 + x1

11 = 1011 = x3 + x1 + 1

19 = 10011 = x4 + x1 + 1

15 = 1111 = x3 + x2 + x1 + 1

29 = 11101 = x4 + x3 + x2 + 1

13 = 1101 = x3 + x2 + 1

35 = 100011 = x5 + x1 + 1

45 = 101101 = x5 + x3 + x2 + 1

# **Лабораторная работа №7**

**Перемежение/деперемежение данных в информационно-вычислительных системах**

**Цель работы**: приобретение практических навыков использования методов перемежения/деперемежения двоичных данных в информационных системах.

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов перемежения/деперемежения двоичных данных в информационных системах.

2. Разработать приложение для реализации метода перемежения/деперемежения символов в сообщениях на основе двоичного алфавита.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**7.1. Теоретические сведения**

**Перемежение** представляет собой такую перестановку символов, при которой стоявшие рядом символы оказываются разделенными несколькими другими символами. Такая процедура предпринимается с целью преобразования групповых ошибок (пакетов ошибок) в одиночные ошибки, с которыми легче бороться с помощью блочного и сверточного кодирования.

Существуют специальные коды, корректирующие пакетные ошибки, однако на практике чаще используют **перемежение/деперемежение** совместно с традиционными кодами. **Пакеты (группы) ошибок** – ошибки, носящие взаимозависимый характер.

Идея **перемежения/деперемежения** состоит в следующем. Если биты каждого кодового слова Хn передаются не в обычной последовательности, а через интервалы, превышающие ожидаемую **длину пакета ошибок** (в промежутки между битами одного слова вставляются биты других кодовых слов), то при возникновении такого типа ошибки обратная перемежению операция – деперемежение – разнесет («размажет») группу ошибок по всей совокупности кодовых слов, составляющих данное сообщение.

**Длина пакета** в нашем случае – это число рядом расположенных ошибочных битов. Например, если Хn = 1011111, а Yn = 1000011, то длина пакета ошибок составляет 3 бита.

Предложено много алгоритмов перемежения/деперемежения. Наиболее простыми являются блочные. При блочном перемежении входные биты делятся на блоки, которые последовательно записываются в строки некоторой таблицы, приведенной для наглядности на рис. 7.1



Передаваемая последовательность (1010110011…) делится на блоки по 5 битов. Каждый блок записывается в отдельную строку таблицы по порядку. Сообщение для передачи или хранения формируется при считывании символов из таблицы по столбцам: 11010000001… .

Деперемежение производится в обратной последовательности. Для данного примера **глубина перемежения** (разница между позициями одного и того же символа до и после перемежения) равна 4: например, второй символ после перемежения станет шестым. Особенностью является неизменная позиция первого символа.

В общем случае **выбор глубины перемежения зависит от двух факторов**. С одной стороны, чем больше расстояние между соседними символами, тем большей длины пакет ошибок может быть исправлен. С другой стороны, чем больше глубина перемежения, тем сложнее аппаратно-программная реализация оборудования и больше задержка сигнала. Для борьбы с длинными пакетами ошибок желательно увеличивать размеры таблицы. Однако это приводит к увеличению задержки в отправке и декодировании сообщения.

**7.2 Практическое задание**

1. Необходимо разработать авторское приложение в соответствии с целью лабораторной работы. По умолчанию используется блочный перемежитель/деперемежитель. По желанию студент может использовать иной. Задание выполняется по указанию преподавателя в соответствии с вариантом из таблицы.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Вариант | Используемый корректирую щий код | Длина пакета ошибок | Число столбцов в матрице | Длина сообщения, байт | Длина информационного слова, *k,* бит |
| 1 | Хемминга | 3, 4, 5 | 5 | 15 | 6 |
| 2 | Циклический | 4, 5, 6 | 5 | 15 | 7 |
| 3 | Хемминга | 3, 4, 5 | 6 | 15 | 5 |
| 4 | Итеративный | 3, 5, 6 | 6 | 15 | 6 |
| 5 | Циклический | 4, 5, 7 | 6 | 15 | 6 |
| 6 | Итеративный | 3, 5, 7 | 7 | 12 | 5 |
| 7 | Хемминга | 3, 5, 7 | 7 | 13 | 6 |
| 8 | Циклический | 3, 6, 7 | 7 | 13 | 7 |
| 9 | Итеративный | 4, 5, 8 | 8 | 13 | 5 |
| 10 | Хемминга | 4, 6, 8 | 7 | 14 | 5 |
| 11 | Циклический | 3, 5, 7 | 7 | 14 | 6 |
| 12 | Итеративный | 3, 4, 8 | 8 | 14 | 7 |
| 13 | Хемминга | 4, 5, 6 | 8 | 15 | 5 |
| 14 | Циклический | 3, 5, 6 | 7 | 15 | 6 |
| 15 | Итеративный | 3, 5, 7 | 7 | 15 | 7 |
| 16 | Хемминга | 3, 6, 7 | 8 | 15 | 6 |

За основу разрабатываемого приложения может быть взято приложение из выполненной лабораторной работы, соответствующей заданному корректирующему коду.

Местоположение заданной группы ошибок выбирается (генерируется) случайным образом. Необходимо для группы ошибок каждой длины сгенерировать 30−40 случайных ситуаций. После деперемежения и исправления ошибок в сообщении сравнить.

**7.3 Выполнение практического задания**

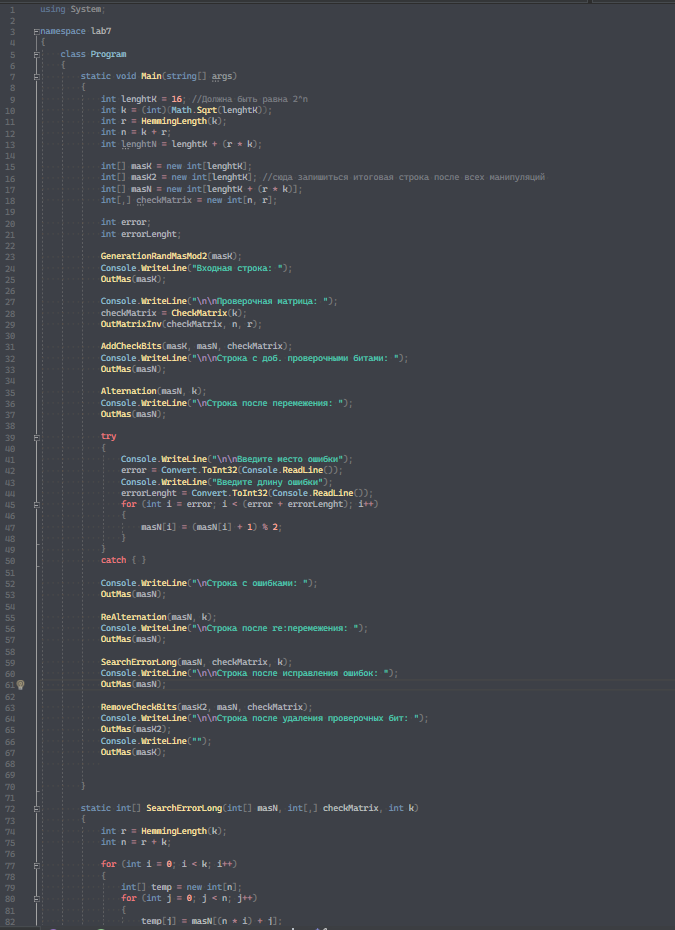


Рис. *7*.2 – Код файла Program.cs



Рис. *7*.3 – Код файла Program.cs

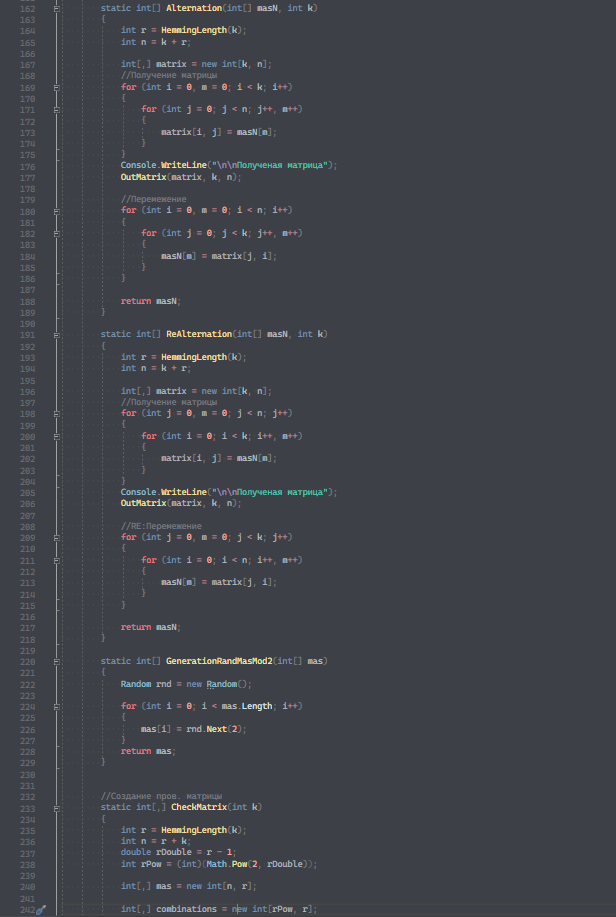


Рис. *7*.4 – Код файла Program.cs

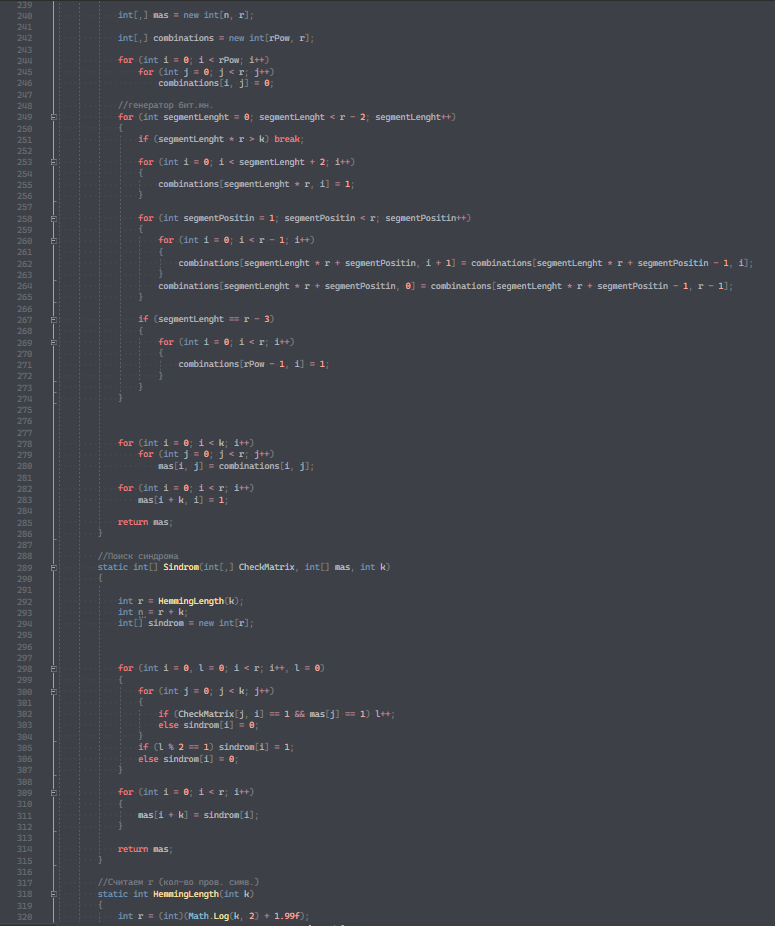


Рис. *7*.5 – Код файла Program.cs



Рис. *7*.6 – Код файла Program.cs

**7.4 Ответы на вопросы**

1. *Пояснить назначение и особенности использования технологии перемежения/деперемежения данных в ИС.*

При такой технологии уменьшается влияние пакетных ошибок. Данные перед передачей по каналу связи, переставляются в заданном порядке, а в приемной части восстанавливается исходный порядок. При этом пакетная ошибка, возникшая в канале связи, превращается в набор рассредоточенных во времени одиночных ошибок, которые проще обнаруживаются и исправляются с помощью кодов, исправляющих ошибки.

1. *Что такое группирующиеся ошибки и с чем, по Вашему мнению, связано их появление в каналах передачи данных, в полупроводниковой памяти, на магнитных носителях?*

Пакеты (группы) ошибок – ошибки, носящие взаимозависимый характер.

Выделяют две основные *причины возникновения ошибок* при передаче информации в сетях:

- сбои в какой-то части оборудования сети или возникновение неблагоприятных объективных событий в сети (например, коллизий при использовании метода случайного доступа в сеть). Как правило, система передачи данных готова к такого рода проявлениям и устраняет их с помощью планово предусмотренных средств;

- помехи, вызванные внешними источниками и атмосферными явлениями. Помехи - это электрические возмущения, возникающие в самой аппаратуре или попадающие в нее извне. Наиболее распространенными являются флуктуационные (случайные) помехи. Они представляют собой последовательность импульсов, имеющих случайную амплитуду и следующих друг за другом через различные промежутки времени. Примерами таких помех могут быть атмосферные и индустриальные помехи, которые обычно проявляются в виде одиночных импульсов малой длительности и большой амплитуды. Возможны и сосредоточенные помехи в виде синусоидальных колебаний. К ним относятся сигналы от посторонних радиостанций, излучения генераторов высокой частоты. Встречаются и смешанные помехи. В приемнике помехи могут настолько ослабить информационный сигнал, что он либо вообще не будет обнаружен, либо искажен так, что “единица” может перейти в “нуль” и наоборот.

1. *Что такое глубина перемежения и как она влияет на эффективность перемежения/деперемежения данных?*

Глубина перемежения - разница между позициями одного и того же символа до и после перемежения.

Чем больше расстояние между соседними символами, тем большей длины пакет ошибок может быть исправлен.

Чем больше глубина перемежения, тем сложнее аппаратно-программная реализация оборудования и больше задержка сигнала. Для борьбы с длинными пакетами ошибок желательно увеличивать размеры таблицы. Однако это приводит к увеличению задержки в отправке и декодировании сообщения.

# **Лабораторная работа №8**

**Сжатие/распаковка данных методом барроуза - уилера**

**Цель работы**: приобретение практических навыков использования метода Барроуза Уилера для сжатия/распаковки данных.

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/ разархивации) данных на основе метода Барроуза Уилера (BurrowsWheeler transform, BWT).

2. Разработать приложение для реализации метода Барроуза Уилера.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**8.1. Теоретические сведения**

Сжатие информации является одним из способов ее кодирования. В основе сжатия данных, как одна из первопричин, лежит **избыточность**.

**Основная цель сжатия** – обеспечить более компактное представление данных, вырабатываемых источником, т. е. уменьшить физический объем сообщений, генерируемых источником, и сократить время его передачи (читай – стоимость) по каналам связи. **Фундаментальная теорема К. Шеннона о кодировании информации** утверждает, что «стоимость кодирования всегда не меньше энтропии источника, хотя может быть сколь угодно близка к ней». Поэтому для любого алгоритма сжатия всегда имеется некоторый предел степени (или эффективности) сжатия, определяемый **энтропией входного потока** (или сжимаемого сообщения).

Все алгоритмы сжатия преобразуют входной поток данных, минимальной единицей которых является бит, а максимальной – байт или несколько байт. Основными техническими характеристиками процессов сжатия и результатов их работы являются: • степень сжатия (англ. compress rating), или отношение R (англ. ratio) объемов исходного (до сжатия, Vдс) и результирующего (после сжатия, Vпс) потоков данных (сообщений); • скорость сжатия − время, затрачиваемое на сжатие некоторого объема информации входного потока до получения из него эквивалентного выходного потока; • качество сжатия − величина, показывающая, насколько сильно сжат выходной поток при помощи применения к нему повторного сжатия по этому же или иному алгоритму.

Степень сжатия R обычно оценивается следующим образом:

R1 = (Vпс / Vдс) · 100%,

R2 = (Vдс − Vпс) / Vдс= (1 − R1) · 100%.

Первое отношение показывает, какую часть объема сообщения (файла) до сжатия занимает сообщение (файл) после сжатия; второе отношение выражает основной физический смысл сжатия и показывает степень сжатия.

Существуют различные подходы к реализации сжатия информации. Они отличаются математической базой, уровнем сложности (простоты) практической реализации, форматом кодируемого потока данных, степенью соответствия сжимаемых и распакованных данных.

По критерию, связанному с характером или форматом данных или степенью соответствия сжимаемых данных распакованным, все методы сжатия разделяют на два класса: обратимое и необратимое сжатие, или иначе: сжатие без потерь и сжатие с частичной потерей информации.

Понятно, что недопустимы никакие потери при упаковке текстовых документов, кодов компьютерных программ, файлов баз данных.

**Метод Барроуза – Уилера.**

**BWT-преобразование (англ. Burrows-Wheeler Transform)** – техника сжатия информации (в особенности текстов), основанная на преобразовании, открытом в 1983 г. BWT не сжимает данные в классическом понимании процесса, но преобразует блок данных в формат, исключительно подходящий для сжатия.

BWT оперирует сразу целым блоком данных, который выделяется из входного потока (сообщения).

**Прямое преобразование (формально – сжатие) выполняется в 4 этапа:**

выделяется блок данных (строка длиной k символов некоторого алфавита мощностью N), который обозначим символом М;

составляется таблица W1 размером k×k всех циклических сдвигов входной строки M: W1 = (M);

производится лексикографическая (в алфавитном порядке) сортировка строк таблицы W1, в результате чего получается таблица W2 того же размера;

в качестве выходной строки (обозначим ее BWT(М), z) выбирается последний столбец (Мk) таблицы W2 преобразования и номер строки z, совпадающей с исходной строкой М. Как видим, выходная строка (сжатое сообщение) всегда по объему превышает входную.

**8.2 Практическое задание**

1. Разработать авторское приложение в соответствии с целью лабораторной работы. Входной блок данных может иметь произвольную длину.

2. С помощью приложения выполнить прямое и обратное преобразования 3 отдельных блоков данных, состоящих:

а) из собственного имени (можно краткий вариант записи);

б) собственной фамилии;

в) варианта в соответствии с таблицей ниже.

Можно использовать любой из известных методов сортировки символов массива.

Выполнить качественный сравнительный анализ длительности процессов прямого и обратного преобразований в зависимости от длины блока данных.

|  |  |
| --- | --- |
| Вариант | Входной блок данных, *М* |
| 1 | летоисчисление |
| 2 | времяпрепровождение |
| 3 | мультимиллионер |
| 4 | семенохранилище |
| 5 | песнетворчество |
| 6 | электрифицированный |
| 7 | водоворотоподобный |
| 8 | самообороноспособность |
| 9 | малосимпатичный |
| 10 | достопримечательность |
| 11 | сорокадневный |
| 12 | телегаммааппарат |
| 13 | полуторапроцентный |
| 14 | гидроаэроионизация |
| 15 | экстраординарный |

3. Перевести первые 3 символа из блока данных, указанного в варианте таблицы, в бинарную последовательность в соответствии с кодами ASCII. Выполнить прямое и обратное преобразование. Оценить время прямого и обратного преобразований.

4. Результаты оформить в виде отчета по установленным правилам.

**8.3 Выполнение практического задания**



Рис. *8.1* – Код файла Program.cs

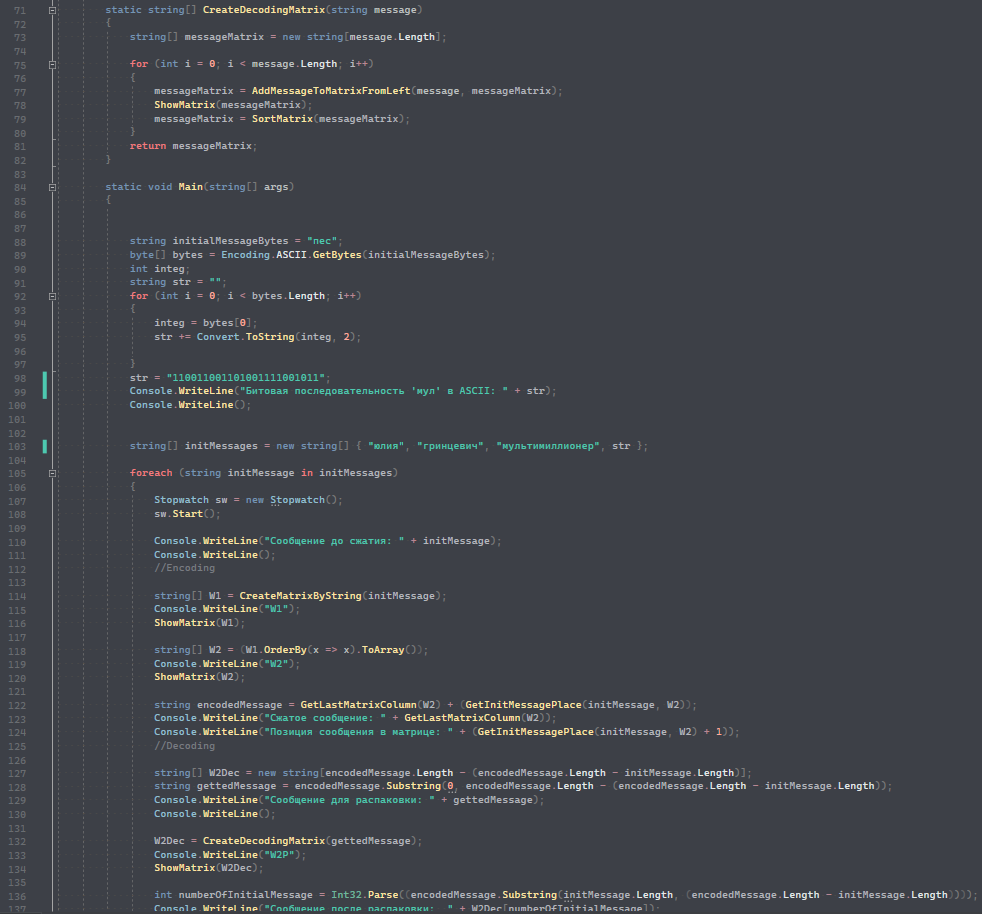


Рис. *8.2* – Код файла Program.cs

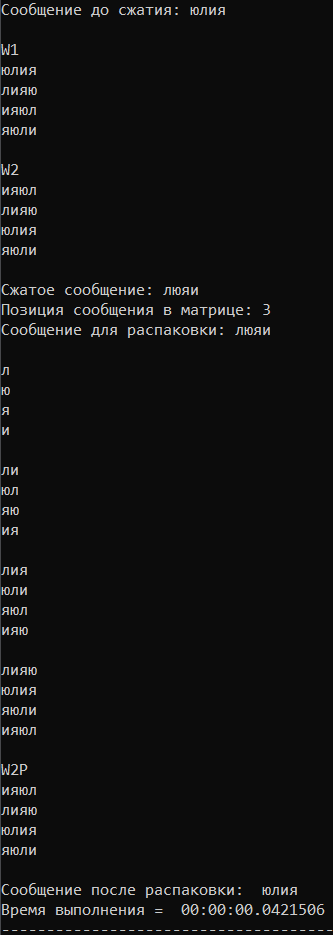
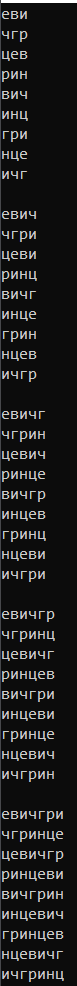
 

Рис. *8.3* – Результат выполнения программы

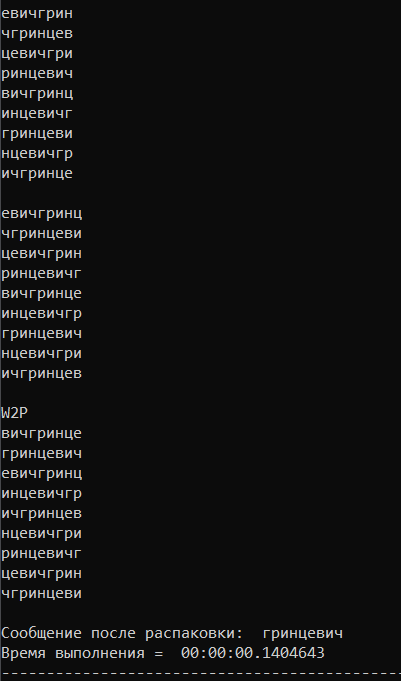


Рис. *8.4* – Результат выполнения программы

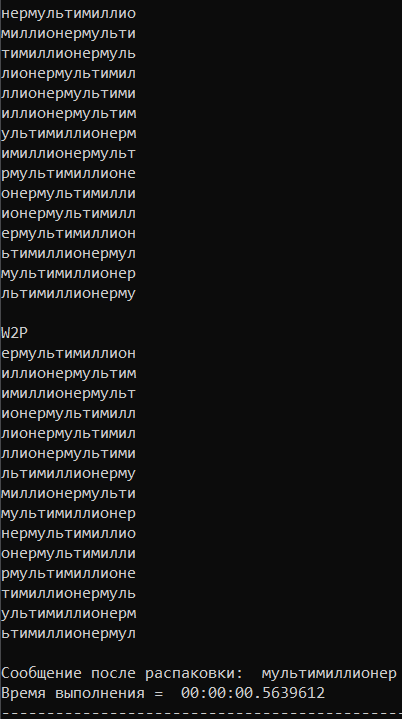
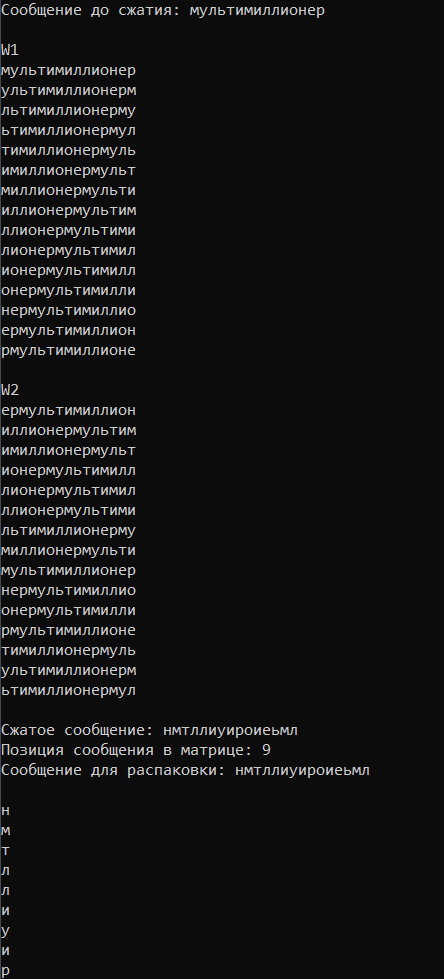


Рис. *8.5* – Результат выполнения программы

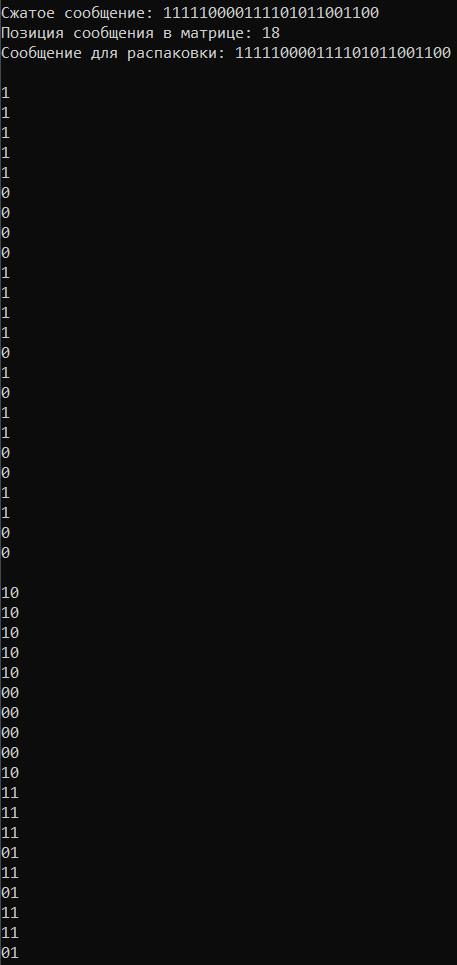
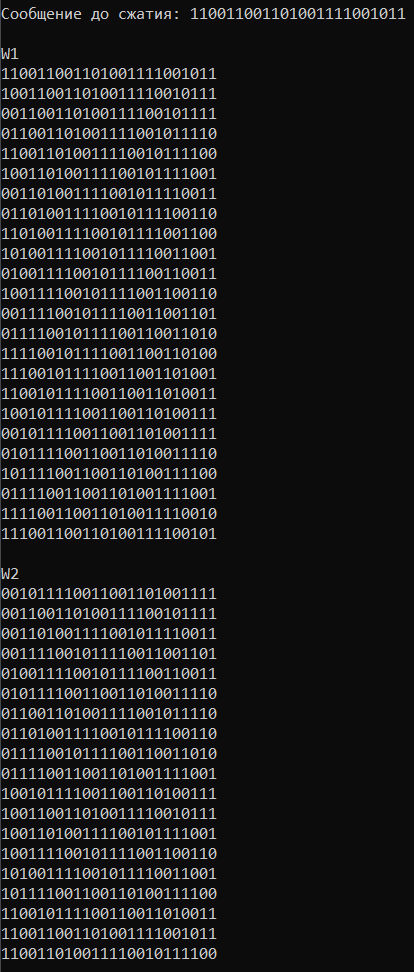


Рис. *8.6* – Результат выполнения программы

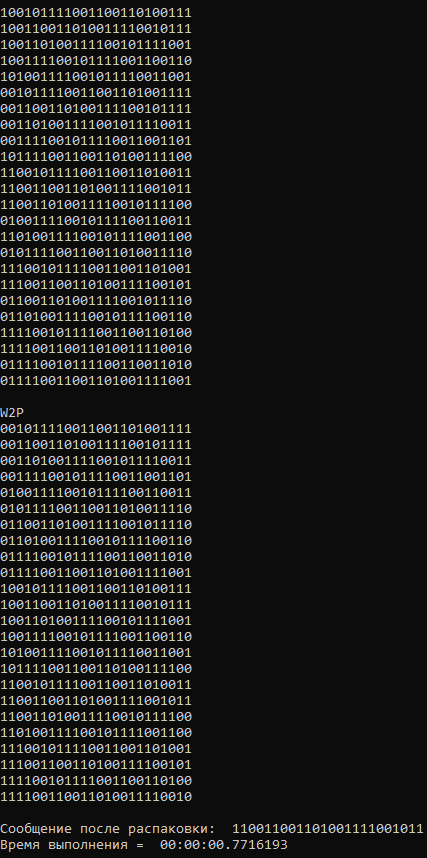


Рис. *8.7* – Результат выполнения программы

**8.4 Ответы на вопросы**

1. **Сформулировать цели применения методов сжатия и архивирования данных.**

Обеспечить более компактное представление данных, вырабатываемых источником, т. е. уменьшить физический объем сообщений, генерируемых источником, и сократить время его передачи (читай – стоимость) по каналам связи. Фундаментальная теорема К. Шеннона о кодировании информации утверждает, что «стоимость кодирования всегда не меньше энтропии источника, хотя может быть сколь угодно близка к ней». Поэтому для любого алгоритма сжатия всегда имеется некоторый предел степени (или эффективности) сжатия, определяемый энтропией входного потока (или сжимаемого сообщения).

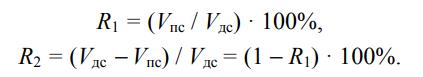
1. **Охарактеризовать основные технические характеристики процессов сжатия/распаковки и результатов.**

• степень сжатия (англ. compress rating), или отношение R (англ. ratio) объемов исходного (до сжатия, Vдс) и результирующего (после сжатия, Vпс) потоков данных (сообщений);

• скорость сжатия − время, затрачиваемое на сжатие некоторого объема информации входного потока до получения из него эквивалентного выходного потока;

• качество сжатия − величина, показывающая, насколько сильно сжат выходной поток при помощи применения к нему повторного сжатия по этому же или иному алгоритму.

1. **Как можно рассчитать степень сжатия файла?**



1. **В каких случаях и почему применяется сжатие без потерь, а в каких – с потерей информации?**

Сжатие с потерей информации реализуется на основе таких известные форматов данных и алгоритмов сжатия, как JPEG и MPEG. Алгоритм JPEG используется при сжатии фотоизображений. Алгоритмы MPEG используют при сжатии видео и музыки. Методы и алгоритмы сжатия с потерей информации применяют обычно для решения так называемых потребительских задач. Это значит, например, что если фотография передается для просмотра, а музыка для воспроизведения, то подобные алгоритмы применять можно. Если же они передаются для дальнейшей обработки, например для редактирования, то никакая потеря информации в исходном файле недопустима. Считают, что на фотографических иллюстрациях, предназначенных для воспроизведения на экране, потеря 5% информации не критична, а в некоторых случаях можно допустить и 20−25% уровень потерь.

1. **В чем сущность символ-ориентированных методов сжатия? Какие известные Вам методы относятся к этому классу?**

Сущность состоит в последовательном анализе сжим. информации с целью поиска повторяющихся или проаналированных ранее в данном документе послед-тей и замене таких послед-тей на более короткие. Методы, основанные на данном подходе, не рассматривают статистические модели, они также не используют коды переменной длины.

1. **Составить алгоритмы сжатия/распаковки данных методом Барроуза − Уилера.**

Прямое преобразование (формально – сжатие) выполняется в 4 этапа:

1) выделяется блок данных (строка длиной k символов некоторого алфавита мощностью N), который обозначим символом М;

2) составляется таблица W1 размером k×k всех циклических сдвигов входной строки M: W1 = (M);

3) производится лексикографическая (в алфавитном порядке) сортировка строк таблицы W1, в результате чего получается таблица W2 того же размера;

4) в качестве выходной строки (обозначим ее BWT(М), z) выбирается последний столбец (Мk) таблицы W2 преобразования и номер строки z, совпадающей с исходной строкой М.

Итак, входной для обратного преобразования является информация вида BWT(М), i. Это преобразование заключается в выполнении k одинаковых шагов, каждый из которых состоит из 2 операций, с целью воссоздания матрицы W2:

1) в крайний справа пустой столбец матрицы записывается последовательность символов Мk;

2) производится лексикографическая сортировка столбцов заполненной части воссоздаваемой матрицы.

1. **Что поменяется, если процедуру формирования матрицы W1 строить на основе циклических сдвигов вправо?**

Ничего.

1. **В каких известных архиваторах используется метод Барроуза − Уилера?**

BWT используется в архиваторе bzip2.

1. **Как вы понимаете рекуррентность (рекурсивность) преобразования по методу Барроуза − Уилера?**

Объект называется рекурсивным, если он содержит сам себя или определен с помощью самого себя. Рекурсия – способ организации вычислительного процесса, при котором процедура или функция в ходе выполнения составляющих её операторов обращается сома к себе.

# **Лабораторная работа №9**

**Сжатие/распаковка данных на основе статистических методов**

**Цель работы**: приобретение практических навыков использования статистических методов Шеннона Фано и Хаффмана (ShannonFano and Huffman coding) для сжатия/распаковки данных..

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/разархивации) данных на основе методов Шеннона Фано и Хаффмана.

2. Разработать приложение для реализации методов Шеннона Фано и Хаффмана.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**9.1. Теоретические сведения**

Статистические алгоритмы позволяют создавать более короткие коды для часто встречающихся и более длинные – для редко встречающихся символов алфавита или конкретного сообщения. В первом случае метод считается статическим статистическим, во втором – динамическим статистическим: вероятностные свойства символов подсчитываются для конкретного сообщения или потока данных.

Частота или вероятность появления того или иного символа алфавита в произвольном сообщении, лежащая в основе алгоритмов, дали название этим алгоритмам и соответствующим методам. Иногда эти методы называют также префиксными. К примеру, если имеется некоторый код, который записывается как Х1 = А1А2, и другой код – Х2 = А1, то говорят, что Х2 является префиксом Х1. Или если Х1 = 1010, а Х2 = 10101100, то Х2 также является префиксом Х1. Таким образом, использование описываемых методов предусматривает создание кодовой таблицы (подобно кодам ASCII или base64). Формально процедура сжатия (прямое преобразование) состоит в подстановке соответствующего бинарного кода вместо символа исходного алфавита и наоборот – при обратном преобразовании. Методы относятся к классу «сжатие без потерь». Различие между двумя рассматриваемыми методами состоит лишь в особенностях формирования таблицы бинарных кодов. При формировании этой таблицы для обоих методов можно воспользоваться статистическими свойствами алфавитов, полученными при выполнении лабораторной работы № 2.

**Код Шеннона – Фано** не является оптимальным (обеспечивает минимальную избыточность) в общем смысле, хотя и дает оптимальные результаты при некоторых распределениях вероятностей. Для одного и того же распределения вероятностей можно построить, вообще говоря, несколько кодов Шеннона – Фано, и все они могут дать различные результаты. Итак, необходимо выполнить следующие действия: 1) подсчитать вероятностные параметры символов алфавита А = {ai} (реализуется статическая версия алгоритма); 2) отсортировать – обычно в порядке убывания (невозрастания, т. е. могут иметь место повторяющиеся значения) вероятностей р(аi); р(аi) – вероятность появления в сжимаемом сообщении на произвольной позиции символа аi алфавита, т. е. создать таблицу символов алфавита, на основе которого генерируется сжимаемое сообщение; 3) каждому символу отсортированного множества поставить в соответствие бинарный код, для чего это множество (таблица) символов делится на две группы таким образом, чтобы каждая из групп имела приблизительно одинаковую суммарную частоту (вероятность). Очевидно, на первом шаге такая суммарная вероятность в каждой из групп должна быть максимально близка к 0,5. Первому из полученных подмножеств устанавливается первый символ бинарного кода: 0, второй − 1 (или наоборот). Для вычисления следующих битов кодов данная процедура повторяется рекурсивно для каждого из полученных на текущем шаге подмножеств, в котором содержится больше одного символа. Получим таблицу, в которой длина кодовых комбинаций меняется от минимального (lmin) до максимального (lmax) значений.

Алгоритм прямого преобразования: необходимо выполнить одну операцию: заменить символы входного сообщения соответствующими бинарными кодами. Алгоритм обратного преобразования: на входе – сообщение в виде бинарной последовательности.

**Метод Хаффмана.**

Метод основан на алгоритме оптимального префиксного кодирования алфавита: исходный алгоритм Хаффмана является оптимальным для посимвольного кодирования с известным входным распределением вероятностей, т. е. для отдельного кодирования несвязанных символов в таком потоке данных. Отличается от метода Шеннона – Фано лишь в части кодирования символов исходного алфавита. В данном случае бинарные коды создаются на основе дерева, ветви которого обозначаются бинарными символами. Бинарным кодом символа исходного алфавита будет последовательность обозначений ветвей дерева от корня до листа, соответствующего этому символу. В основе бинарного кода лежит следующее положение.

**Для любого заданного алфавита (источника) с N > 2 символами существует оптимальный двоичный код, в котором два наименее вероятных символа (слова) имеют одну и ту же длину и отличаются лишь последним битом**

Построение дерева начинается с сортирования символов исходного алфавита в порядке убывания (невозрастания). Далее выбираются два символа (ai, aj) с наименьшими вероятностями (р(ai), р(aj)) и объединяются в узел. Ветви этого узла обозначаются «1» и «0». Этот узел рассматривается далее как новый, виртуальный символ (aij), которому будет соответствовать вероятность р(aij) = р(ai) + р(aj). Такой виртуальный символ будет рассматриваться далее наравне с остальными символами исходного алфавита. Два его потомка из дальнейшего рассмотрения исключаются. Создаются новые узлы дерева по тому же принципу. Корень дерева образуют два символа с наибольшими вероятностями.

**9.2 Практическое задание**

1. Разработать авторское приложение в соответствии с целью лабораторной работы.
2. С помощью приложения выполнить прямое и обратное преобразования сообщения, состоящего из собственных имени и фамилии. Можно использовать любой из известных методов сортировки символов массива. Метод кодировки (Шеннона − Фано, Хаффмана) использовать по указанию преподавателя.

При этом таблица отсортированных символов строится:

а) на основе данных, полученных в лабораторной работе № 2;

б) динамически, на основе анализа сжимаемого сообщения.

3. Определить эффективность (в сравнении с кодами ASCII) сжатия сообщения.

4. Результаты оформить в виде отчета по установленным правилам

**9.3 Выполнение практического задания**

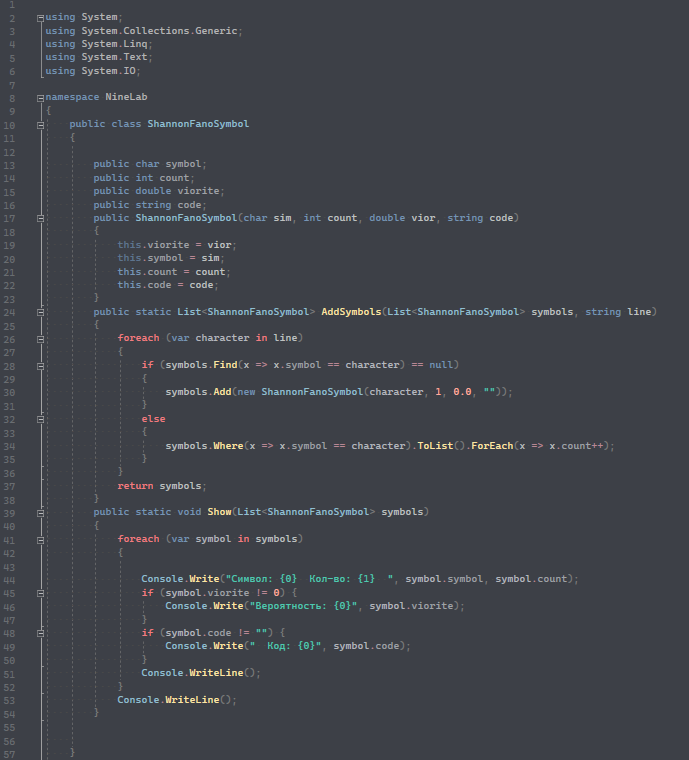


Рис. *9.1* – Код файла Program.cs

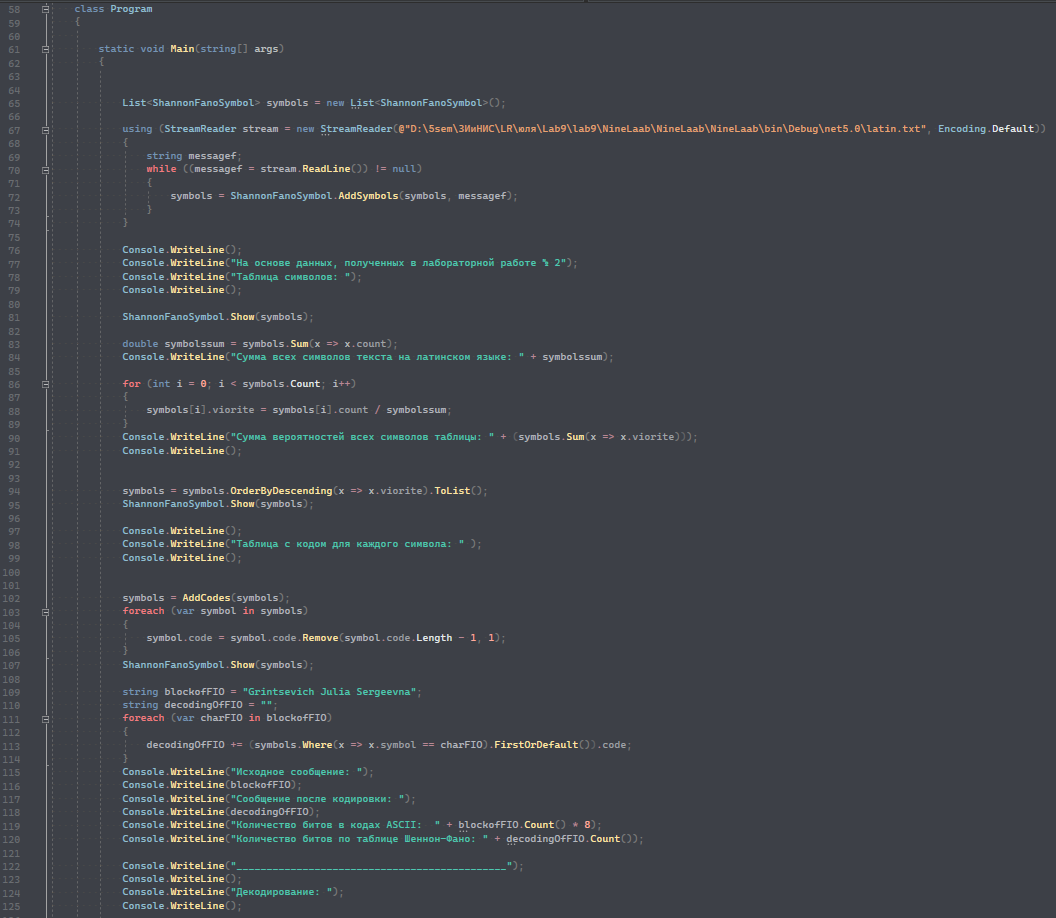


Рис. *9.2* – Код файла Program.cs

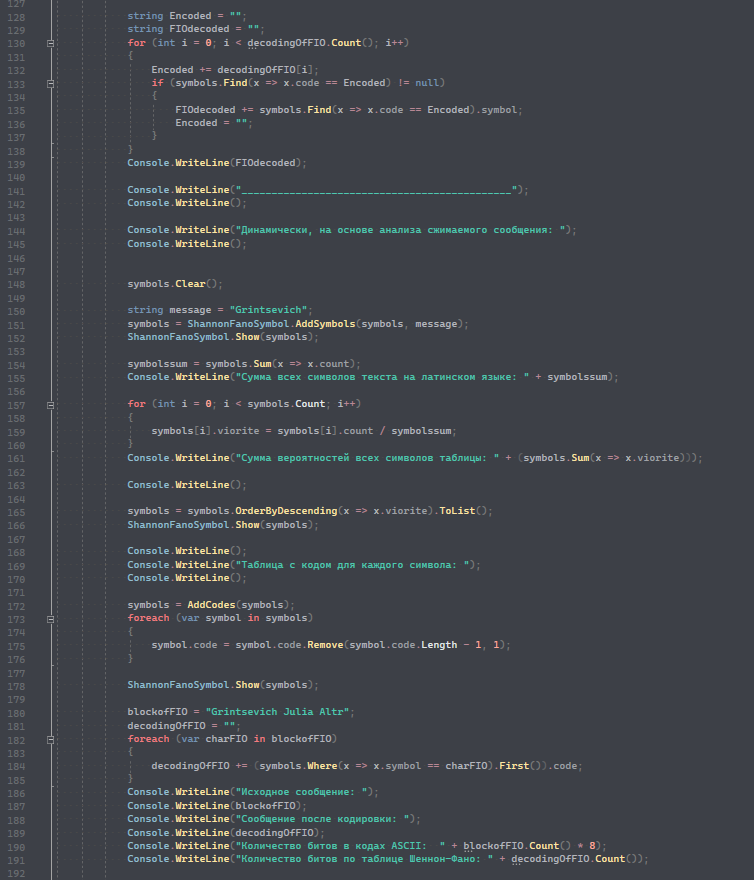


Рис. *9.3* – Код файла Program.cs



Рис. *9.4* – Код файла Program.cs

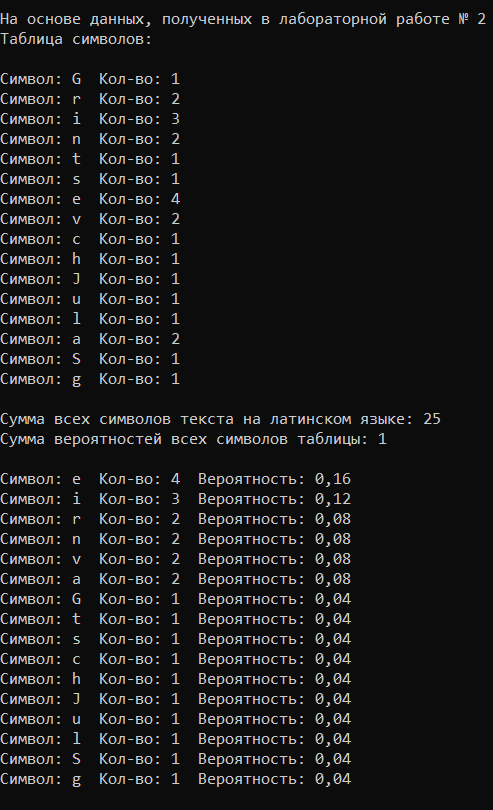


Рис. *9.5* – Результат выполнения программы

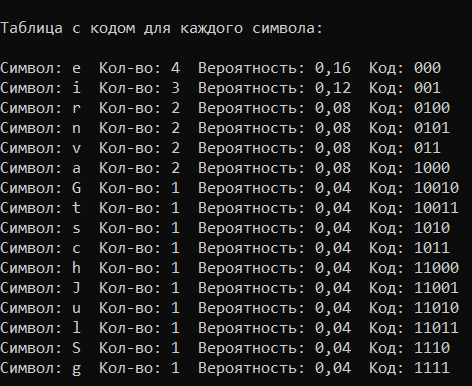


Рис. *9.6* – Результат выполнения программы

**9.4 Ответы на вопросы**

1. *Что такое бинарное дерево, чем характеризуется его структура?*

Бинарное дерево-это корневое дерево, которое также является упорядоченным деревом (оно же плоское дерево), в котором каждый узел имеет не более двух дочерних элементов. Корневое дерево естественным образом дает понятие уровней (расстояние от корня), поэтому для каждого узла понятие дочерних элементов может быть определено как узлы, соединенные с ним уровнем ниже.

1. *Какие коды называются префиксными?*

|  |
| --- |
| Префиксный код (англ. prefix code) — код, в котором никакое кодовое слово не является префиксом какого-то другого кодового слова. |

Предпочтение префиксным кодам отдается из-за того, что они упрощают декодирование. Поскольку никакое кодовое слово не выступает в роли префикса другого, кодовое слово, с которого начинается файл, определяется однозначно, как и все последующие кодовые слова.

# **Лабораторная работа №10**

**Сжатие/распаковка данных методов Лемпеля-Зива**

**Цель работы**: приобретение практических навыков использования статистических методов Шеннона Фано и Хаффмана (ShannonFano and Huffman coding) для сжатия/распаковки данных..

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию методов сжатия/распаковки (архивации/разархивации) данных на основе методов Шеннона Фано и Хаффмана.

2. Разработать приложение для реализации методов Шеннона Фано и Хаффмана.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента.

**10.1. Теоретические сведения**

В 1977 г. Авраам Лемпель и Якоб Зив выдвинули идею формирования «словаря» общих последовательностей анализируемых (сжимаемых) данных. При этом сжатие данных осуществляется за счет замены записей соответствующими кодами из словаря. Классический алгоритм Лемпеля − Зива – LZ77, названный так по году представления метода, формулируется следующим образом: **«если в проанализированном (сжатом) ранее выходном потоке уже встречалась подобная последовательность байт, причем запись о ее длине и смещении от текущей позиции короче, чем сама эта последовательность, то в выходной файл записывается ссылка (смещение, длина), а не сама последовательность».**

Известный метод сжатия RLE, который заключается в записи вместо последовательности одинаковых символов одного символа и их количества, является подклассом LZ77.

**Суть метода LZ77** (как и последующих его модификаций) состоит в следующем: **упаковщик постоянно хранит некоторое количество последних обработанных символов в буфере**. По мере обработки входного потока вновь поступившие символы попадают в конец буфера, сдвигая предшествующие символы и вытесняя самые старые. Размеры этого буфера, называемого также **скользящим словарем**, варьируются в разных реализациях систем сжатия. Скользящее окно имеет длину **n**, т. е. в него помещается **n** символов, и состоит из двух частей:

последовательности длины **n1 = n − n2** уже закодированных символов (словарь);

упреждающего буфера (буфера предварительного просмотра, lookahead) длиной n2 – буфера кодирования.

Пусть к текущему моменту времени закодировано t символов: S1, S2, ..., St. Тогда словарем будут являться n1 предшествующих символов: St − (n1 − 1), St − (n1 − 1)+1, …, St. В буфере находятся ожидающие кодирования (сжатия) символы St+1, St+2, …, St+n2. Если n2 ≥ t, то словарем будет являться вся уже обработанная часть входной последовательности

**Нужно найти самое длинное совпадение между строкой буфера кодирования, начинающейся с символа St + 1, и всеми фразами словаря.**

Эти фразы могут начинаться с любого символа St − (n1 − 1), St − (n1 − 1) + 1, …, St, выходить за пределы словаря, вторгаясь в область буфера, но должны лежать в окне. Буфер не может сравниваться сам с собой. Длина совпадения не должна превышать размера буфера. Полученная в результате поиска фраза St − (р − 1), St − (р − 1) + 1, St − (р − 1) + (q − 1) кодируется с помощью двух чисел:

смещения (англ. offset) от начала буфера p;

длины соответствия, или совпадения (англ. match length) q.

Ссылки (p и q − указатели) однозначно определяют фразу. Дополнительно в выходной поток записывается символ s, следующий за совпавшей строкой буфера.

Длина кодовой комбинации (триады – p, q, s) на каждом шаге определяется соотношением

l(с ) = logN n1 +logN n2 + 1, (10.1)

где N – мощность алфавита.

После каждого шага окно смещается на q + 1 символов вправо и осуществляется переход к новому циклу кодирования. Величина сдвига объясняется тем, что мы реально закодировали именно q + 1 символов: q – с помощью указателя и 1 − с помощью тривиального копирования.

**10.2 Практическое задание**

1. Разработать авторское приложение в соответствии с целью лабораторной работы.
2. С помощью приложения выполнить прямое и обратное преобразования сообщения, состоящего из собственных имени и фамилии. Можно использовать любой из известных методов сортировки символов массива. Метод кодировки (Шеннона − Фано, Хаффмана) использовать по указанию преподавателя.

При этом таблица отсортированных символов строится:

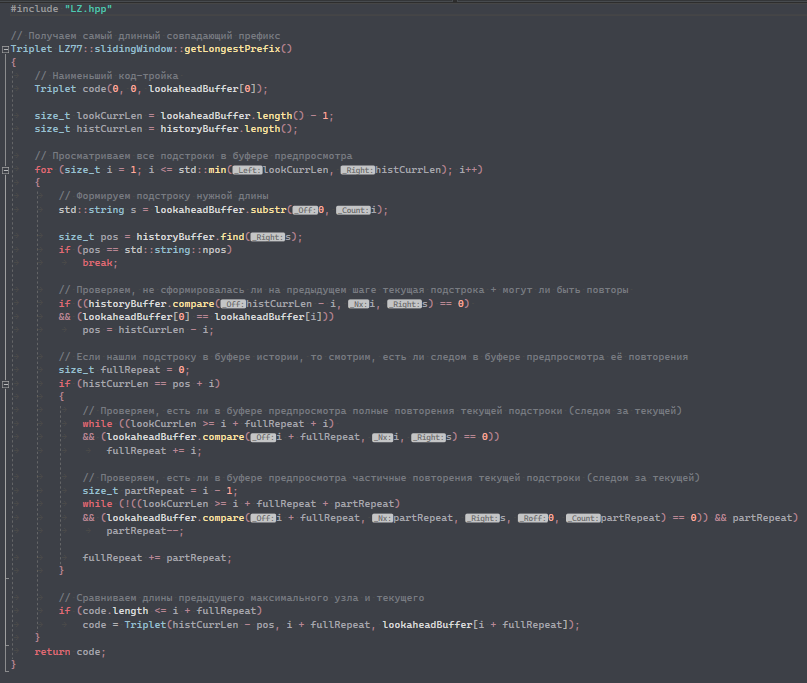
а) на основе данных, полученных в лабораторной работе № 2;

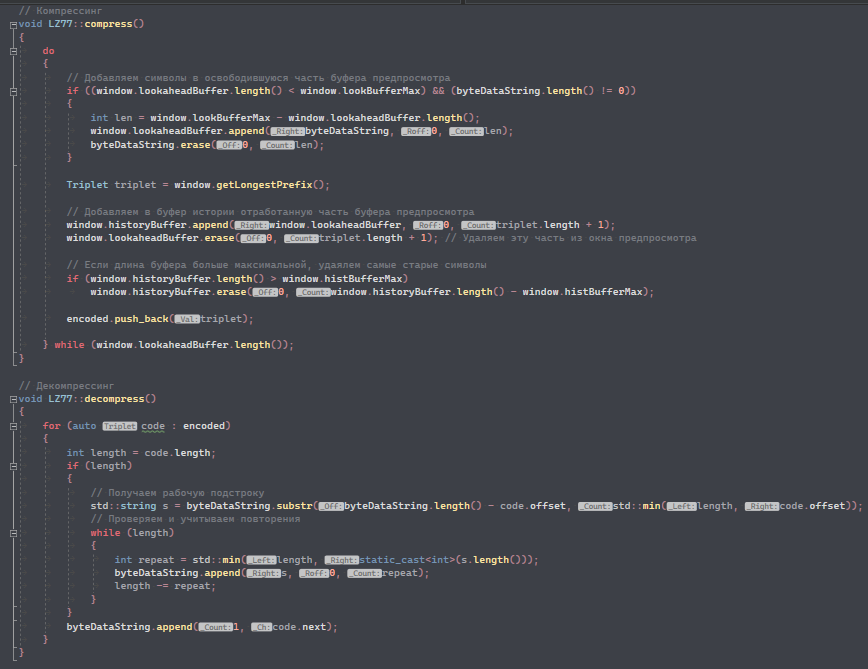
б) динамически, на основе анализа сжимаемого сообщения.

3. Определить эффективность (в сравнении с кодами ASCII) сжатия сообщения.

4. Результаты оформить в виде отчета по установленным правилам

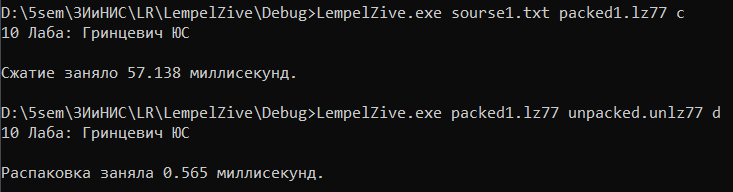
**10.3 Выполнение практического задания**

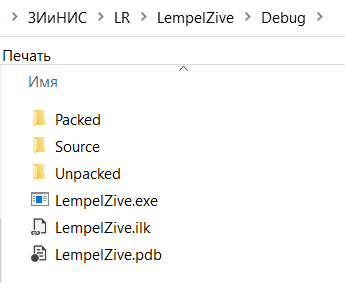
****

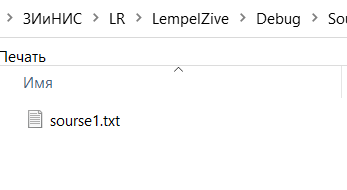
****

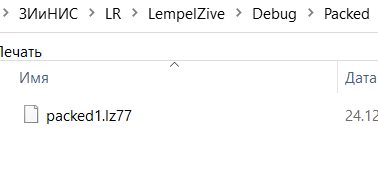
****

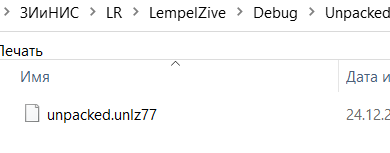
****

****

****

****

****

****

**10.4 Ответы на вопросы**

*1. К какому классу методов сжатия относится метод LZ и почему?*

**LZ** **методы** **относятся** к **классу** адаптивных словарных кодов. При использовании таких кодов составляется адаптивная таблица, или словарь, в которую записываются последовательности символов источника и соответствующие им кодовые слова.

*2. Какое значение имеет последний символ в сжатом текстовом сообщении методом LZ?*

Показывает каким символом прервалась выбранная последовательность символов

*3. Какие модификации метода LZ77 Вам известны? В чем заключается особенность модификаций?*

алгоритмы типа **LZ77** и алгоритмы типа LZ78

Суть метода LZ77 (как и последующих его модификаций) состоит в следующем: упаковщик постоянно хранит некоторое количество последних обработанных символов в *буфере*. По мере обработки входного потока вновь поступившие символы попадают в конец буфера, сдвигая предшествующие символы и вытесняя самые старые

# **Лабораторная работа №11**

**Сжатие/распаковка данных арифметическим методом**

**Цель работы**: приобретение практических навыков использования  
арифметических методов сжатия/распаковки данных.

**Задачи:**

1. Закрепить теоретические знания по алгебраическому описанию и использованию арифметических методов сжатия/распаковки (архивации/разархивации) данных.

2. Разработать приложение для реализации арифметических методов.

3. Результаты выполнения лабораторной работы оформить в виде описания разработанного приложения, методики выполнения экспериментов с использованием приложения и результатов эксперимента

**11.1. Теоретические сведения**

Пpи аpифметическом сжатии (кодиpовании) текст пpедставляется вещественными числами в интеpвале от 0 до 1. По меpе анализа текста отобpажающий его интеpвал уменьшается, а количество битов для его пpедставления возpастает. Очеpедные символы текста сокpащают величину интеpвала, исходя из значений соответствующих веpоятностей.

**Основная идея арифметического метода сжатия заключается в том, чтобы присваивать коды не отдельным символам, а их последовательностям.**

Таким образом, как и во всех энтропийных алгоритмах, исходной является информация о частоте встречаемости каждого символа алфавита. Алгоритмы прямого и обратного преобразований базируются на операциях с «рабочим отрезком».

Рабочим отрезком называется интервал [a; b] с расположенными на нем точками. Причем точки расположены таким образом, что длины образованных ими отрезков пропорциональны (или равны) частоте (вероятности) появления соответствующих символов.

**Прямое преобразование (сжатие).** Один шаг сжатия (кодирования) заключается в простой операции: берется кодируемый символ, для него ищется соответствующий участок на рабочем отрезке. Найденный участок становится новым рабочим отрезком. Его тоже необходимо разбить с помощью точек.

Это и последующие разбиения отрезка (на шаге i) подразумевают определение новых значений верхней (Hi) и нижней (Li) границ для всего участка и осуществляются по следующим правилам:

**Hi(αj)= Li − 1 + (Hi − 1 – Li − 1) · H(αj)0;**

**Li(αj)= Li − 1 + (Hi − 1 – Li − 1) · L(αj)0,**

где αj – j-й символ сжимаемой последовательности, Li − 1 и Hi − 1 – соответственно нижняя и верхняя границы рабочего отрезка на (i − 1)-м шаге, L(αj)0 и H(αj)0 – соответственно исходные нижняя и верхняя границы символа αj.

Результатом кодирования цепочки символов является любое число с итогового рабочего отрезка. Обычно таким числом является нижняя граница указанного отрезка. Таким образом, итогом сжатия входной последовательности будет число.

**Обратное преобразование (декомпрессия).** Для восстановления исходного сообщения необходима информация:

о значении числа, являющегося итогом сжатия сообщения;

количестве символов в сжатом сообщении;

вероятностных параметрах всех символов исходного сообщения (таблица вероятностей).

Как и при сжатии, вначале необходимо начальный рабочий отрезок [0; 1) разбить на интервалы, длины которых равны вероятностям появления соответствующих символов, т. е. создать рабочий отрезок, полностью соответствующий первоначальному отрезку.

На каждом шаге обратного преобразования выбираем отрезок, в который попадает текущее число (код). Символ, который соответствует данному отрезку, является очередным символом восстановленного (распакованного) сообщения.

В общем случае код символа, восстанавливаемого на шаге i, вычисляется соотношением:

**код i = [код (i − 1) − L(αi − 1)0] / [H(αi − 1)0 − L(αi − 1)0],**

где код (i − 1) – число, анализ которого производился на предыдущем шаге – (i − 1)-м; H(αi − 1)0 и L(αi − 1)0 – соответственно верхняя и нижняя исходные границы символа сообщения, восстановленного на предыдущем шаге.

**11.2 Практическое задание**

1. Разработать авторское приложение в соответствии с целью лабораторной работы.

2. С помощью приложения выполнить прямое и обратное преобразования сообщений в соответствии с таблицей.

Каждый студент выполняет задание, состоящее из двух частей. Первая часть предусматривает кодирование/декодирование сообщения, указанного в 2-м столбце, вторая часть – составного сообщения, полученного конкатенацией последовательностей из 2-го столбца, указанных в 3-м столбце.

3. Дать оценку возможности переполнения при выполнении вычислений.

4. Сравнить характеристики арифметического сжатия с вероятностными алгоритмами.

****

**11.3 Выполнение практического задания**

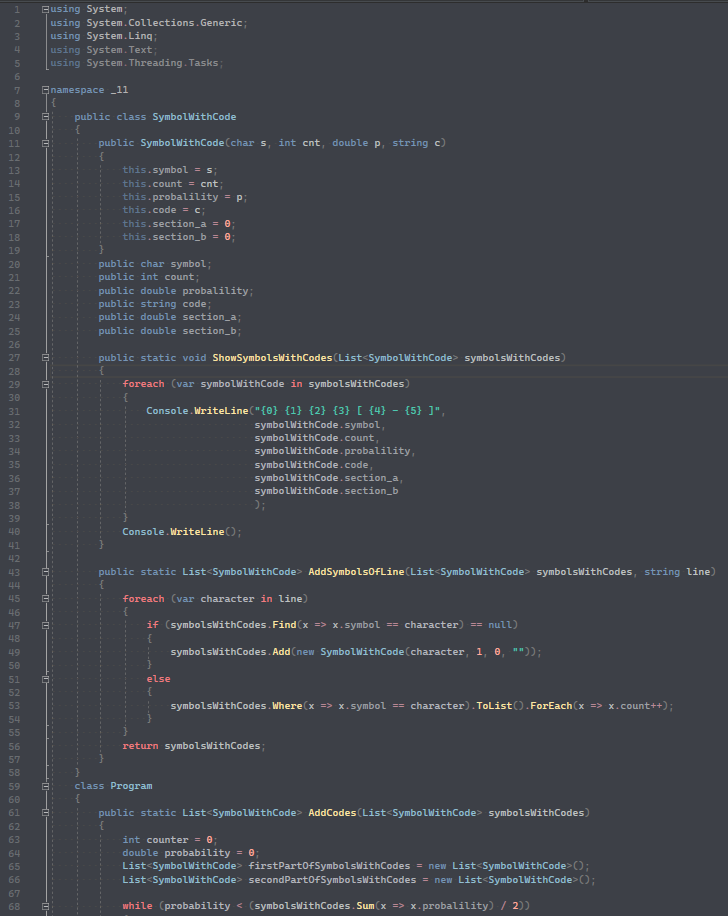
****

Рис. 10*.1* – Код файла Program.cs

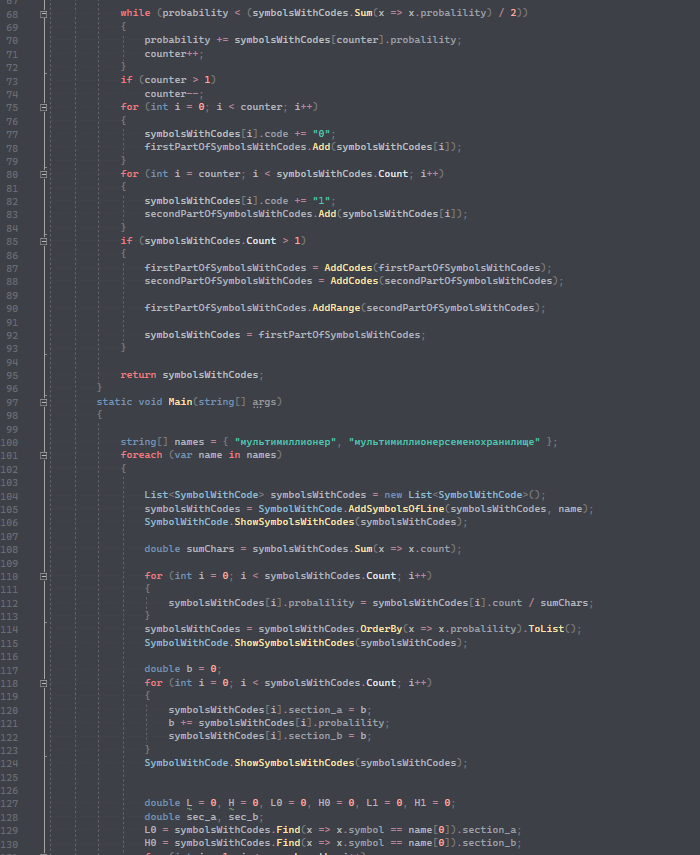
****

Рис. *10.2* – Код файла Program.cs

****

Рис. *10.3* – Код файла Program.cs

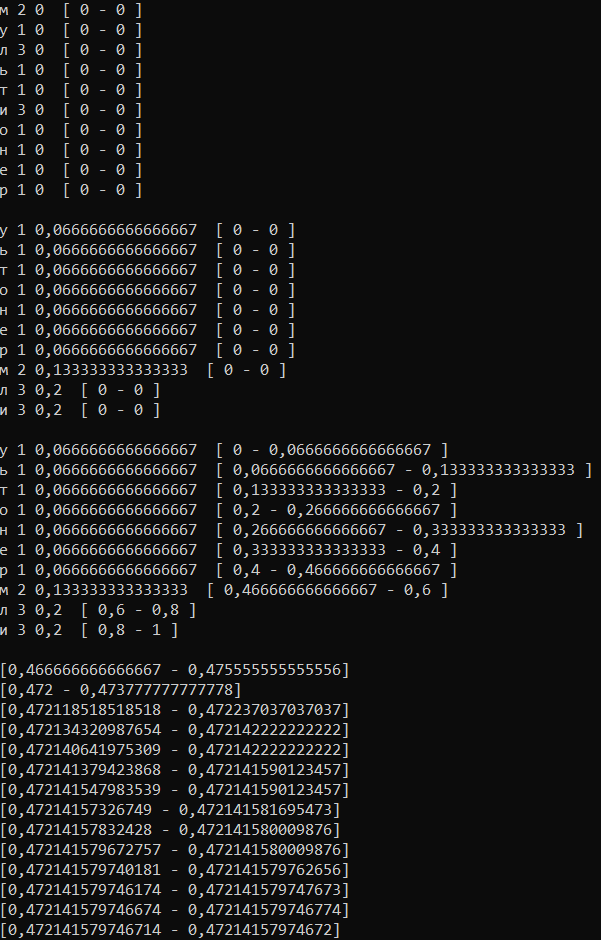
****

Рис. *10.4* – Результат

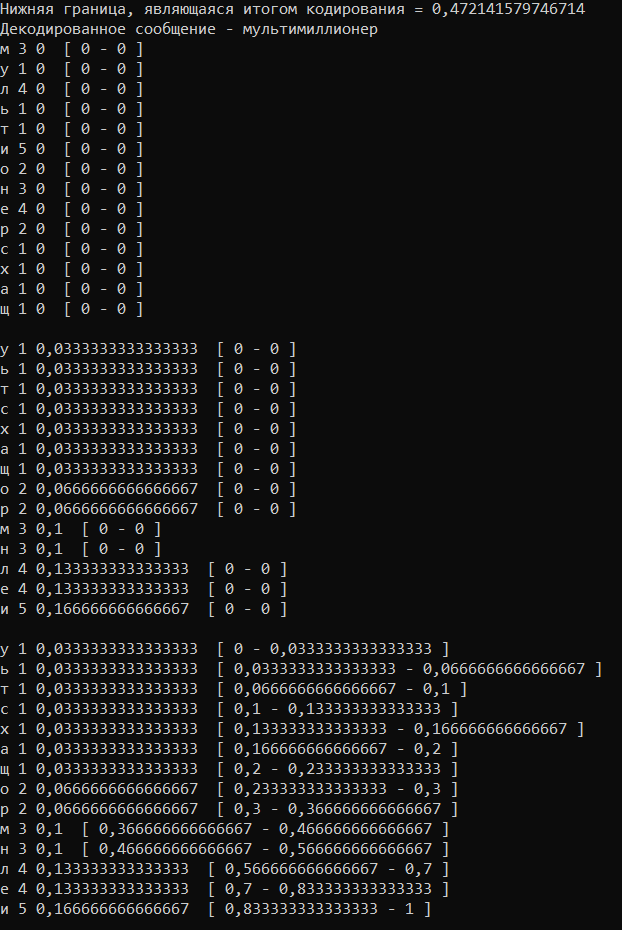
****

Рис. *10.5* – Результат

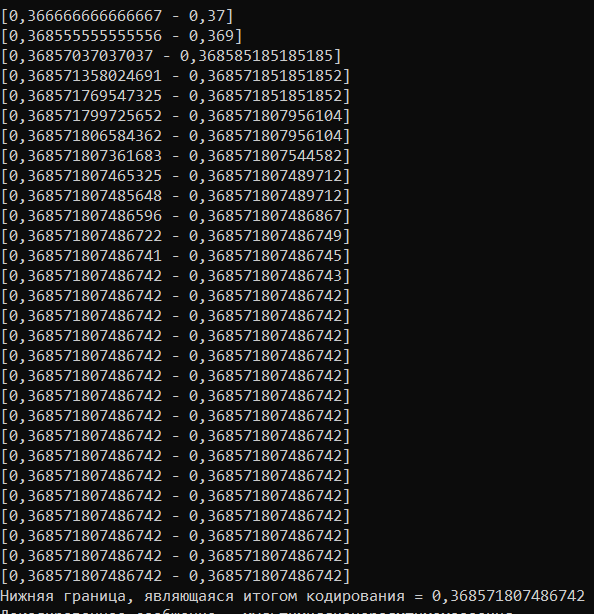
****

Рис. *10.6* – Результат

**11.4 Ответы на вопросы**

1. В чем заключается суть арифметического метода сжатия?

Основная идея арифметического метода сжатия заключается в том, чтобы присваивать коды не отдельным символам, а их последовательностям

2. Пояснить принцип определения границ отрезков на каждом шаге (на каждом рабочем интервале) прямого и обратного преобразования.

Алгоритмы прямого и обратного преобразований базируются на операциях с «рабочим отрезком». Рабочим отрезком называется интервал [a; b] с расположенными на нем точками. Причем точки расположены таким образом, что длины образованных ими отрезков пропорциональны (или равны) частоте (вероятности) появления соответствующих символов

3. Какой недостаток, свойственный префиксным методам, отсутствует в арифметическом кодировании?

Требуемая для представления интервала [Hi; Li] точность возрастает вместе с длиной анализируемого текста; постепенное выполнение алгоритма помогает преодолеть эту проблему.

# **Лабораторная работа №12**

**Основы теории чисел и их использование в криптографии**

**Цель работы**: приобретение практических навыков выполнения операций с числами для решения задач в области криптографии и разработка приложений для автоматизации этих операций.

**12.1. Теоретические сведения**

В основе современной криптографии лежит теория чисел.

Теория чисел, или высшая арифметика, – раздел математики, изучающий натуральные числа и иные похожие величины.

**Определение 1.** Множество всех целых чисел (обозначим буквой Z) есть набор всех действительных чисел без дробной части: {..., –3, –2, –1, 0, 1, 2, 3, ...}.

**Определение 2.** Натуральные числа являются подмножеством целых чисел и образуют множество N: {1, 2, 3, ...}.

**Определение 3**. Делимость – одно из основных понятий теории чисел. Если для некоторого целого числа a и натурального числа b существует целое число q, при котором bq = a, то говорят, что число a делится на b. В этом случае b называется делителем числа a, а a называется кратным числу b. При этом используются следующие обозначения:

a ⋮ b – a делится на b, или b | a – b делит a.

Из последнего определения следует, что:

• любое натуральное число является делителем нуля;

• единица является делителем любого целого числа;

• любое натуральное число является делителем самого себя.

**Определение 4.** Делитель a называется собственным делителем числа b, если 1 < |a| < |b|, и несобственным – в противном случае.

**Определение 5.** Всякое целое число а можно представить с помощью положительного целого числа b равенством вида а = bq + r, 0 ≤ r ≤ b. Число q называется неполным частным, а число r – остатком отделения а на b.

**Если число не имеет делителей, кроме самого себя и единицы, то оно называется простым, а если у числа есть еще делители, то составным.**

**Определение 6.** Натуральное число n называется простым, если n > 1 и не имеет положительных делителей, отличных от 1 и n.

Простое число не делится без остатка ни на одно другое число.

**Основная теорема арифметики**. Всякое натуральное число n, кроме 1, можно представить как произведение простых множителей: n = p1p2p3...pz, z > 1.

**решето Эратосфена –** алгоритм для поиска простых чисел.

**Простое число** – число, которое делится на цело только на себя и единицу.

Есть массив чисел, и мы рассматриваем первый элемент. Далее вычеркиваем все числа, которые нацело делятся на этот элемент.

**Алгоритм Евклида – используется для нахождения НОД двух целых положительных чисел.**

**Расширенный** **алгоритм** **Евклида** — **это** **расширение** **алгоритма** **Евклида**, которое вычисляет кроме наибольшего общего делителя (НОД) целых чисел a и b ещё и коэффициенты соотношения Безу, то есть целые x и y

**12.2 Практическое задание**

Используя L\_PROST, найти все простые числа в интервале [2, n]. Значение n соответствует варианту из табл. 1.2, указанному преподавателем. Подсчитать количество простых чисел в указанном интервале. Сравнить это число с n/ln(n) (см. выше пример 15).

Повторить п. 1 для интервала [m, n]. Сравнить полученные результаты с «ручными» вычислениями, используя «решето Эратосфена»

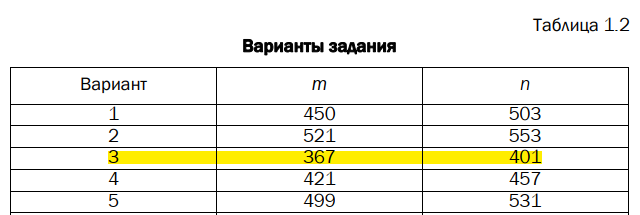
Записать числа m и n в виде произведения простых множителей (форма записи – каноническая).

Проверить, является ли число, состоящее из конкатенации цифр m ǀǀ n (табл. 1.2), простым.

Найти НОД (m, n).

Разработать авторское приложение в соответствии с целью лабораторной работы. Приложение должно реализовывать следующие операции: • вычислять НОД двух либо трех чисел; • выполнять поиск простых чисел.

С помощью созданного приложения выполнить задания по условиям п. 1 и 2.

****

**12.3 Выполнение практического задания**

****

Рис. 12*.1* – Код файла Program.cs



Рис. 12*.2* – Код файла Program.cs



Рис. 12*.3* – Код файла Program.cs

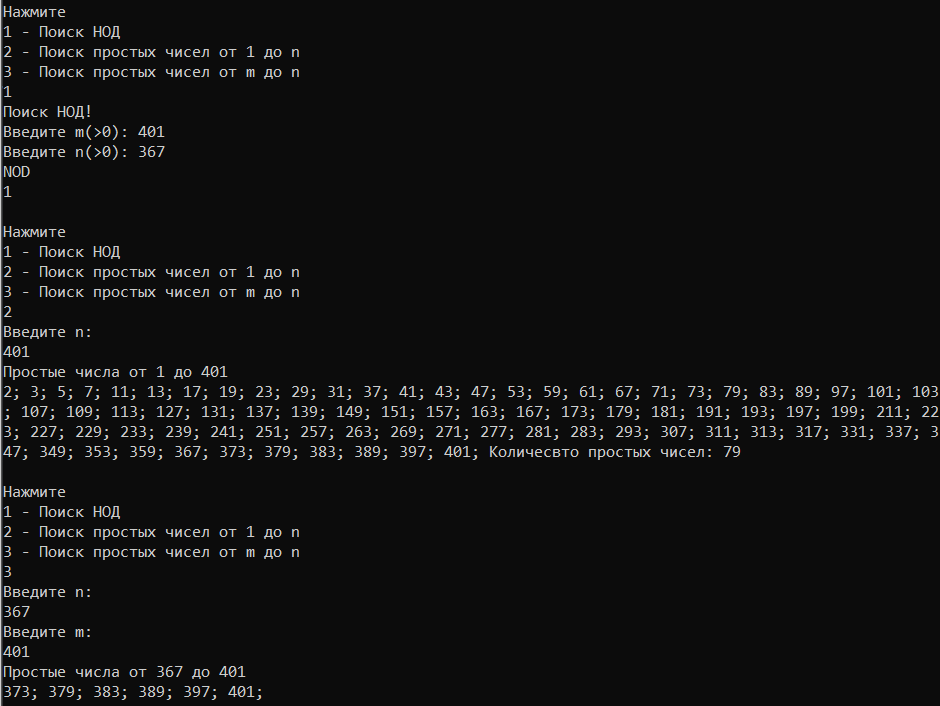


Рис. *12.4* – Результат

**12.4 Ответы на вопросы**

*1. Дать определение понятий: целое число, натуральное число, делимость чисел, собственный делитель, НОД.*

Целое число (от латинского integer, означающего "целое")[a] в разговорной речи определяется как число, которое может быть записано без дробной составляющей.

Натуральные числа являются подмножеством целых чисел и образуют множество N: {1, 2, 3, ...}.

Делимость – одно из основных понятий теории чисел. Если для некоторого целого числа a и натурального числа b существует целое число q, при котором bq = a, то говорят, что число a делится на b. В этом случае b называется делителем числа a, а a называется кратным числу b.

Делитель a называется собственным делителем числа b, если 1 < |a| < |b|, и несобственным – в противном случае.

Наибольшее целое число, которое делит без остатка числа a и b, называется наибольшим общим делителем этих чисел – НОД (a, b).

*2. Сформулировать основную теорему арифметики. Представить примеры ее применения.*

Каждое натуральное число *n*, за исключением единицы, раскладывается в произведение простых сомножителей, причем единственным образом с точностью до порядка следования сомножителей:

https://konspekta.net/lektsiiorgimg/baza13/674373247956.files/image005.gif ,

где *p*1, …, *pk* – простые числа.

*3. Пояснить сущность проблемы факторизации и ее связь с прикладной криптографией.*

Сложность решения задачи разложения больших чисел на простые сомножители, известной как «проблема фактори-зации», определяет криптостойкость некоторых алгоритмов асимметричной криптографии, в частности алгоритма RSA. Свойство 4. Любое четное число, большее 2, представимо в виде суммы двух простых чисел, а любое нечетное, большее 5, представимо в виде суммы трех простых чисел.