Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»

Факультет физики и информационных технологий
Кафедра автоматизированных систем обработки информации

СОГЛАСОВАНО
Заведующий кафедрой
автоматизированных систем
обработки информации
А.В.Воруев
__ 18.04. 2023 г.

СОГЛАСОВАНО
Декан
факультета физики и
информационных технологий
Д.Л.Коваленко
18.04. 2023 г.

ЭЛЕКТРОННЫЙ УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

для специальности
1-53 01 02 Автоматизированные системы обработки информации

составители: старший преподаватель кафедры АСОИ Кулинченко В.Н.
д.т.н., профессор кафедры АСОИ Демиденко О.М.
старший преподаватель кафедры АСОИ Кучеров А.И.

Рассмотрено и утверждено
на заседании кафедры АСОИ
__18 апреля___ 2023 г., протокол № _9_

Рассмотрено и утверждено
на заседании научно-методического
совета университета
__29 апреля___ 2023 г., протокол № _8_

Гомель 2023

1 ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Электронный учебно-методический комплекс (ЭУМК) по дисциплине «Информационные системы и технологии» представляет собой комплекс систематизированных учебных, методических и вспомогательных материалов, предназначенных для использования в образовательном процессе специальности 1-53 01 02 – Автоматизированные системы обработки информации.

ЭУМК разработан в соответствии со следующими нормативными документами:

1. Положением об учебно-методическом комплексе на уровне высшего образования, утвержденном постановлением Министерства образования Республики Беларусь от 08.11.2022 № 427.

2. Образовательным стандартом ОСВО 1-53 01 02-2021 г. и учебным планом ГГУ имени Ф.Скорины регистрационный № I 53-1-21/УП, дата утверждения 31.05.2021.

3. Учебной программой по учебной дисциплине «Информационные системы и технологии» для специальности 1-53 01 02 Автоматизированные системы обработки информации, утвержденной 28.07.2021, регистрационный номер УД-2021-115/уч.

Цель создания ЭУМК – обеспечить приобретение теоретических знаний и практических навыков в области управления и обработки информации, овладение студентами основами современных информационных технологий передачи, преобразования и хранения информации.

ЭУМК направлен на оказание помощи студентам в овладении теоретическими основами и практическими навыками обработки информации в различных ИС и АСУП с использованием современных облачных технологий и сервисов. Организация изучения дисциплины специализации на основе ЭУМК предполагает продуктивную образовательную деятельность, позволяющую сформировать социально-личностные и профессиональные компетенции будущих специалистов, обеспечить развитие познавательных и созидательных способностей личности.

ЭУМК способствует успешному осуществлению учебной деятельности, дает возможность планировать и осуществлять самостоятельную управляемую работу студентов, обеспечивает рациональное распределение учебного времени по темам учебной дисциплины и совершенствование методики проведения занятий.

ЭУМК состоит из теоретического, практического и вспомогательного разделов. Теоретический раздел содержит тексты лекций по разделам передачи и преобразования информации курса основ информационных технологий. Практический раздел содержит методические рекомендации к лабораторным и практическим работам, тестовые задания и вопросы для

самоконтроля. Вспомогательный раздел содержит учебную программу и список литературы.

Теоретический раздел содержит лекционный материал по темам учебной программы. В разделе так же содержатся рекомендации по организации и выполнению управляемой самостоятельной работы.

Практический раздел включает в себя темы лабораторных занятий и задания с краткими методическими указаниями по выполнению лабораторных работ. В разделе так же приводятся некоторый набор тестовых заданий и для примера указана часть вопросов для самоконтроля.

Вспомогательный раздел содержит необходимые элементы учебно-программной документации по дисциплине с указанием рекомендуемой литературы (основной, дополнительной).

Все разделы ЭУМК соответствуют содержанию учебной программы и объему учебного плана.

ЭУМК по дисциплине государственного компонента «Информационные системы и технологии» изучается студентами 1 курса дневной формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации»; студентами 1 курса заочной формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации»; студентами 1 курса заочной интегрированной со средним специальным образованием формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации».

Дневная форма обучения: всего часов по плану-108, аудиторное количество часов – 108; из них: лекционных занятий – 32 (в том числе УСР 6), лабораторных работ – 24.

Форма отчётности – экзамен в 1 семестре.

Заочная форма обучения: всего часов по плану-108, аудиторное количество часов – 14, из них: лекционных занятий –8, лабораторных работ – 6.

Форма отчётности – экзамен во 2 семестре.

Заочная форма обучения (интегрированная на основе среднего специального образования): всего часов по плану-108, аудиторное количество часов – 14, из них: лекционных занятий – 8, лабораторных работ – 6.

Форма отчётности – экзамен во 2 семестре.

2 ТЕКСТЫ ЛЕКЦИЙ

ВВЕДЕНИЕ

Вещественные, энергетические и информационные процессы в современном обществе. Понятие «технология». Основные аспекты понятия технология. Информационная технология. Особенности понятия информационной технологии. Современные информационные технологии и их особенности. Информационная технология и информационная система. Основные виды информационных систем.

Технология — это комплекс научных и инженерных знаний, реализованных в приемах труда, наборах материальных, технических, энергетических, трудовых факторов производства, способах их соединения для создания продукта или услуги, отвечающих определенным требованиям. Поэтому технология неразрывно связана с машинизацией производственного или непроизводственного, прежде всего управленческого процесса. Управленческие технологии основываются на применении компьютеров и телекоммуникационной техники.

Согласно определению, принятому ЮНЕСКО, информационная технология — это комплекс взаимосвязанных, научных, технологических, инженерных дисциплин, изучающих методы эффективной организации труда людей, занятых обработкой и хранением информации; вычислительную технику и методы организации и взаимодействия с людьми и производственным оборудованием, их практические приложения, а также связанные со всем этим социальные, экономические и культурные проблемы. Сами информационные технологии требуют сложной подготовки, больших первоначальных затрат и наукоемкой техники. Их введение должно начинаться с создания математического обеспечения, формирования информационных потоков в системах подготовки специалистов.

Термин информационная система (ИС) может использоваться как в широком, так и в узком смысле.

В широком смысле *информационная система* – совокупность технического, программного и организационного обеспечения, а также персонала, предназначенная для того, чтобы своевременно обеспечивать надлежащих людей надлежащей информацией. Также одно из наиболее широких определений ИС дал специалист в области баз данных и информационных систем, ведущий научный сотрудник ИПР РАН М.Р. Когаловский: «*информационной системой* называется комплекс, включающий вычислительное и коммуникационное оборудование, программное обеспечение, лингвистические средства и информационные ресурсы, а также системный персонал и обеспечивающий поддержку динамической информационной модели некоторой части реального мира для удовлетворения информационных потребностей пользователей».

В узком смысле информационной системой называют только подмножество компонентов ИС в широком смысле, включающее базы данных, СУБД и специализированные прикладные программы. ИС в узком смысле рассматривают как программно-аппаратную систему, предназначенную для автоматизации целенаправленной деятельности

конечных пользователей, обеспечивающую, в соответствии с заложенной в нее логикой обработки, возможность получения, модификации и хранения информации.

В любом случае основной задачей ИС является удовлетворение конкретных информационных потребностей в рамках конкретной предметной области. Современные ИС де-факто немыслимы без использования баз данных и СУБД, поэтому термин «информационная система» на практике сливается по смыслу с термином «система баз данных».

В идеале в рамках предприятия должна функционировать единая корпоративная информационная система, удовлетворяющая все существующие информационные потребности всех сотрудников, служб и подразделений. Однако на практике создание такой всеобъемлющей ИС слишком затруднено или даже невозможно, вследствие чего на предприятии обычно функционируют несколько различных ИС, решающих отдельные группы задач: управление производством, финансово-хозяйственная деятельность и т.д. Часть задач бывает «покрыта» одновременно несколькими ИС, часть задач – вовсе не автоматизирована. Такая ситуация получила название «лоскутной автоматизации» и является довольно типичной для многих предприятий.

Три процесса в информационной системе производят информацию, в которой нуждаются организации для принятия решений, управления, анализа проблем и создания новых изделий или услуг – это ввод, обработка и вывод. В процессе ввода фиксируются или собираются непроверенные сведения внутри организации или из внешнего окружения. В процессе обработки этот сырой материал преобразуется в более значимую форму. На стадии вывода обработанные данные передаются персоналу или процессам, где они будут использоваться. Информационные системы также нуждаются в обратной связи, которая является возвращаемыми обработанными данными, нужными для того, чтобы приспособить элементы организации для помощи в оценке или исправлении обработанных данных.

РАЗДЕЛ 1. МЕТОДОЛОГИЧЕСКИЙ БАЗИС ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Взаимосвязь информационной технологии как научной дисциплины с другими научными направлениями. Понятие системы. Основные свойства системы. Структура, архитектура и цель системы. Основные аспекты структуры сложной системы. Разработка архитектуры информационной системы.

Классификации информационных систем.

Классификация по архитектуре.

По степени распределённости отличают:

- настольные (desktop), или локальные ИС, в которых все компоненты (БД, СУБД, клиентские приложения) находятся на одном компьютере;

распределённые (distributed) ИС, в которых компоненты распределены по нескольким компьютерам.

Распределённые ИС, в свою очередь, разделяют на:

- файл-серверные ИС (ИС с архитектурой «файл-сервер»);
- клиент-серверные ИС (ИС с архитектурой «клиент-сервер»).

В файл-серверных ИС база данных находится на файловом сервере, а СУБД и клиентские приложения находятся на рабочих станциях.

В клиент-серверных ИС база данных и СУБД находятся на сервере, а на рабочих станциях находятся клиентские приложения.

В свою очередь, клиент-серверные ИС разделяют на двухзвенные и многозвенные.

В двухзвенных (англ. two-tier) ИС всего два типа «звеньев»: сервер баз данных, на котором находятся БД и СУБД (back-end), и рабочие станции, на которых находятся клиентские приложения (front-end). Клиентские приложения обращаются к СУБД напрямую.

В многозвенных (англ. multi-tier) ИС добавляются промежуточные «звенья»: серверы приложений (application servers). Пользовательские клиентские приложения не обращаются к СУБД напрямую, они взаимодействуют с промежуточными звеньями. Типичный пример применения многозвенности – современные веб-приложения, использующие базы данных. В таких приложениях помимо звена СУБД и клиентского звена, выполняющегося в веб-браузере, имеется как минимум одно промежуточное звено – веб-сервер с соответствующим серверным ПО.

Классификация по степени автоматизации. По степени автоматизации ИС делятся на:

- автоматизированные: информационные системы, в которых автоматизация может быть неполной (то есть требуется постоянное вмешательство персонала);
- автоматические: информационные системы, в которых автоматизация является полной, то есть вмешательство персонала не требуется или требуется только эпизодически.

«Ручные ИС» («без компьютера») существовать не могут, поскольку существующие определения предписывают обязательное наличие в составе ИС аппаратно-программных средств. Вследствие этого понятия «автоматизированная информационная система», «компьютерная информационная система» и просто «информационная система» являются синонимами.

Классификация по характеру обработки данных. По характеру обработки данных ИС делятся на:

- информационно-справочные, или информационно-поисковые ИС, в которых нет сложных алгоритмов обработки данных, а целью системы является поиск и выдача информации в удобном виде;
- ИС обработки данных, или решающие ИС, в которых данные подвергаются обработке по сложным алгоритмам. К таким системам в первую очередь относят автоматизированные системы управления и системы поддержки принятия решений.

Классификация по сфере применения. Поскольку ИС создаются для удовлетворения информационных потребностей в рамках конкретной предметной

области, то каждой предметной области (сфере применения) соответствует свой тип ИС. Перечислять все эти типы не имеет смысла, так как количество предметных областей велико, но можно указать в качестве примера следующие типы ИС: экономическая информационная система, медицинская информационная система, географическая информационная система и т.д.

Классификация по охвату задач (масштабности):

- персональная ИС предназначена для решения некоторого круга задач одного человека.
- групповая ИС ориентирована на коллективное использование информации членами рабочей группы или подразделения.
- корпоративная ИС в идеале охватывает все информационные процессы целого предприятия, достигая их полной согласованности, безызбыточности и прозрачности. Такие системы иногда называют системами комплексной автоматизации предприятия.

Качество системы. Стандарты ISO серии 9000. Назначение и особенности стандартов серии 9000. Модель «уровней зрелости» CMM. Основные модели CMM. Стандарт CMMI. Основные характеристики уровней CMM.

РАЗДЕЛ 2. КОНЦЕПТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Концепция открытых систем. Методологические основы открытых систем. Основополагающие документы, определяющие концепцию открытых систем.

Эталонная модель OSI. Основные понятия модели OSI. Уровни OSI. Основные задачи и выполняемые функции. Понятие стека протоколов.

Общие сведения о стандартах в области информационных технологий. Роль стандартов в области информационных технологий. Уровни и виды стандартов.

Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека. Методология современных информационных систем. Информационные технологии корпоративных и государственных учреждений. Системы класса ERP. Корпоративные порталы. CALS-технологии.

ERP-системы

Система класса ERP (Enterprise Resource Planning - Управление ресурсами предприятия) - это корпоративная информационная система для автоматизации планирования, учета, контроля и анализа всех основных бизнес-процессов и решения бизнес задач в масштабе предприятия (организации). ERP-система помогает интегрировать все отделы и функции компании в единую систему, при этом все департаменты работают с единой базой данных и им проще обмениваться между собой разного рода информацией.

Обычно ERP система включает в себя различные функциональные модули, например, бухгалтерский и налоговый учет, управление складом, транспортировками, казначейство, кадровый учет, управление взаимоотношениями с клиентами. Различные программные модули единой системы ERP позволяют заменить устаревшие разрозненные информационные системы по управлению логистикой, финансами, складом, проектами. Вся информация хранится в единой базе данных, откуда она может быть в любое время получена по запросу.

Внедрение ERP системы – достаточно сложный и длительный процесс. Интеграция ERP системы в бизнес-процессы компании предполагает серьезные изменение логики внутренних процедур в компании, реинжиниринга бизнес-процессов а также значительные изменения в работе ее сотрудников. В связи со сложностью проекта сроки внедрения систем класса ERP достаточно большие (2-3 года). Но внедрение системы класса ERP дает следующие возможности:

- планировать потребности в материалах и комплектующих, сроки и объёмы поставок для выполнения плана производства продукции;
- регулировать наличие продукции (излишки, дефицит) и снижать издержки на ее хранение;
- регулировать процесс производства своевременно реагируя на изменение спроса;
- оптимизировать бизнес-процессы в компании путем сокращения материальных и временных затрат;
- контролировать поставки и качество сервиса для клиентов.

Положительные стороны внедрения ERP системы на предприятии (в организации):

- сокращение уровня страховых запасов;
- своевременность пополнения материально-технических ресурсов;

- повышение оборачиваемости оборотных средств;
- сокращение неликвидных запасов и числа неплановых закупок;
- повышение объемов производства и повышение эффективности
- эффективный контроль расхода материалов;
- повышение эффективности ценообразования;
- снижение трудозатрат на формирование бухгалтерской отчетности;

Система SAP ERP, комплекс решений SAP, Microsoft Dynamics AX и др.

CRM-система (Customer Relationship Management - Управление отношениями с клиентами) - корпоративная информационная система, незаменимый современный инструмент для ведения бизнеса. Дает возможность не просто автоматизировать взаимодействие с клиентами и процесс продаж, а выстроить их работу таким образом, чтобы получать максимальный результат.

Возможности CRM-систем:

- Быстрый доступ к актуальной информации о клиентах;
- Оперативность обслуживания клиентов и проведения сделок;
- Формализация схем взаимодействия с клиентами, автоматизация документооборота;
- Быстрое получение всех необходимых отчетных данных и аналитической информации;
- Снижение операционных затрат менеджеров;
- Контроль работы менеджеров;
- Согласованное взаимодействие между сотрудниками и подразделениями.

Области применения CRM-систем:

CRM система применима в любом бизнесе, где клиент персонифицирован, где высока конкуренция и успех зависит от предоставления наиболее выгодных для клиента условий. Максимального эффекта от внедрения CRM-систем добиваются компании, работающие в областях:

- Услуг;
- Производства;
- Оптовой и розничной торговли;
- Страхования и финансов;
- Телекоммуникации и транспорта;
- Строительства.

Выбор CRM-системы

Проект внедрения CRM-системы обычно связан с глубокими организационными изменениями в компании. В первую очередь, необходимо продумать клиентоориентированную стратегию компании, затем приступать к выбору подходящей для вас CRM-системы.

Основные критерии выбора CRM-системы для управления отношениями клиентами:

- Соответствие функциональных возможностей системы целям бизнеса и стратегии компании;
- Возможность интеграции с другими корпоративными информационными системами;
- Возможность доработки CRM-системы с ориентацией на потребности компании;
- Соответствие CRM техническим требованиям;
- Совокупная стоимость владения CRM-системы (стоимость лицензий, внедрение, сопровождение)
- Доступность услуг по внедрению и подддержке в вашем регионе

CRM система может быть либо самостоятельным программным продуктом, либо входить в состав ERP-системы как модуль (например CRM модуль в ERP-системе Microsoft Dynamics AX).

BI, EIS, DSS, электронный бизнес и коммерция

За последние 10 лет менялись названия и содержание информационно-аналитических систем от информационных систем руководителя (executive information systems, EIS) до систем поддержки принятия решений (decision support systems, DSS) и сейчас до систем бизнес-интеллекта.

Во времена больших ЭВМ и миникомпьютеров, когда у большинства пользователей не было прямого доступа к компьютерам, организации зависели от своих подразделений ИТ, которые обеспечивали их стандартными и параметрическими отчетами. Но чтобы получить отчеты, отличные от стандартных, пользователям нужно было заказывать их разработку и ждать в течение нескольких дней или недель.

Приложения EIS были настроены на нужды руководителей и менеджеров и давали возможность получать основную агрегированную информацию о состоянии их бизнеса в виде таблиц или диаграмм. Обычно они включали регламентные запросы с набором параметров. Такие пакеты обычно разрабатывались силами своих подразделений ИТ. Для получения дополнительной информации и проведения дальнейшего анализа применялись другие приложения или создавались по заказу запросы или отчеты на SQL.

Приложения DSS первого поколения были пакетами прикладных программ с динамической генерацией SQL-скриптов по типу запрашиваемой пользователем информации. Они позволяли аналитикам получать информацию из реляционных БД, не требуя знания SQL. В отличие от EIS приложения DSS могут отвечать на широкий спектр вопросов бизнеса, имеют несколько вариантов представления отчетов и определенные возможности форматирования. Однако гибкость таких пакетов все же была ограничена из-за ориентации на конкретный набор задач.

С приходом ПК и локальных сетей следующее поколение приложений DSS строится уже на основе BI и позволяет пользователю-непрограммисту легко и оперативно извлекать информацию из различных источников,

формировать собственные настраиваемые отчеты или графические представления, проводить многомерный анализ данных. Развитие систем бизнес-интеллекта прошло путь от «толстых» клиентов до Web-приложений, в которых пользователь ведет исследование с помощью браузера и может работать удаленно. Можно также создавать сценарии «что если» и коллективно просматривать и обновлять информацию.

Хотя пользователи корпоративной BI-информации традиционно находятся внутри предприятия, с распространением Web для электронного бизнеса, B2B, CRM и SCM BI-пользователи могут быть и внешними по отношению к предприятию, а в B2C, C2B и на торговых площадках пользователями BI являются пользователи Internet.

BI и хранилища данных. Концепция, методы и средства хранилища данных (Data warehousing) определяют подходы и обеспечивают интеграцию, очистку, ретроспективное хранение информации, предназначенной для анализа , отвечают на вопрос «Как подготовить информацию для анализа?». Технология бизнес-интеллекта определяет методы и средства доступа и оперативного анализа информации в терминах предметной области. BI-средства не обязательно должны работать в инфраструктуре хранилища данных, но в этом случае проблема очистки и согласования данных возлагается на них, причем осуществлять эти операции придется на лету или же предварительно, но для обособленного информационного ресурса. Кроме того, есть эффект влияния на производительность и надежность оперативной системы обработки транзакций. Вот почему хорошей корпоративной практикой является выделение транзакционной и аналитической составляющих и применение для второй различных решений по хранилищу данных. Основные стыки идут не только на уровне информации, но и на уровне метаданных. В случае хранилища данных можно обеспечить централизованное управление метаданными.

Следует отметить, что часто термином «хранилище данных» обозначают систему поддержки принятия решений DSS или информационно-аналитическую систему, основанные на технологиях хранилища данных и бизнес-интеллекта.

Классификация продуктов business intelligence

Сегодня категории BI-продуктов включают: BI-инструменты и BI-приложения. Первые, в свою очередь, делятся на: генераторы запросов и отчетов; развитые BI-инструменты, - прежде всего инструменты оперативной аналитической обработки (online analytical processing, OLAP); корпоративные BI-наборы (enterprise BI suites, EBIS); BI-платформы. Главная часть BI-инструментов делится на корпоративные BI-наборы и BI-платформы. Средства генерации запросов и отчетов в большой степени поглощаются и замещаются корпоративными BI-наборами. Многомерные OLAP-механизмы или серверы, а также реляционные OLAP-механизмы являются BI-инструментами и инфраструктурой для BI-платформ. Большинство BI-инструментов применяются конечными пользователями для доступа, анализа

и генерации отчетов по данным, которые чаще всего располагаются в хранилище, витринах данных или оперативных складах данных. Разработчики приложений используют BI-платформы для создания и внедрения BI-приложений, которые не рассматриваются как BI-инструменты. Примером BI-приложения является информационная система руководителя EIS.

Инструменты генерации запросов и отчетов. Генераторы запросов и отчетов - типично «настольные» инструменты, предоставляющие пользователям доступ к базам данных, выполняющие некоторый анализ и формирующие отчеты. Запросы могут быть как незапланированными (ad hoc), так и иметь регламентный характер. Имеются системы генерации отчетов (как правило, серверные), которые поддерживают регламентные запросы и отчеты. Настольные генераторы запросов и отчетов расширены также некоторыми облегченными возможностями OLAP. Развитые инструменты этой категории объединяют в себе возможности пакетной генерации регламентных отчетов и настольных генераторов запросов, рассылки отчетов и их оперативного обновления, образуя так называемую корпоративную отчетность (corporate reporting) . В ее арсенал входят сервер отчетов, средства рассылки, публикации отчетов на Web, механизм извещения о событиях или отклонениях (alerts). Характерные представители - Crystal Reports, Cognos Impromptu и Actuate e.Reporting Suite.

Корпоративные BI-наборы. EBIS - естественный путь для предоставления BI-инструментов, которые ранее поставлялись в виде разрозненных продуктов. Эти наборы интегрируются в наборы инструментов генерации запросов, отчетов и OLAP. Корпоративные BI-наборы должны иметь масштабируемость и распространяться не только на внутренних пользователей, но и на ключевых заказчиков, поставщиков и др. Продукты BI-наборов должны помогать администраторам при внедрении и управлении BI без добавления новых ресурсов. Из-за тесного родства Web и корпоративных BI-наборов некоторые поставщики описывают свои BI-наборы как BI-порталы. Эти портальные предложения обеспечивают подмножество возможностей EBIS с помощью Web-браузера, однако поставщики постоянно увеличивают их функциональность, приближая ее к возможностям инструментов для «толстых» клиентов. Типичные EBIS поставляют Business Objects и Cognos.

BI-платформы. BI-платформы предлагают наборы инструментов для создания, внедрения, поддержки и сопровождения BI-приложений. Имеются насыщенные данными приложения с «заказными» интерфейсами конечного пользователя, организованные вокруг специфических бизнес-проблем, с целевым анализом и моделями. BI-платформы, хотя и не так быстро растут и широко используются как EBIS, являются важным сегментом благодаря ожидаемому и уже происходящему росту BI-приложений. Стараниями поставщиков реляционных СУБД, создающих OLAP-расширения своих СУБД, многие поставщики платформ, которые предоставили многомерные

СУБД для OLAP, чтобы выжить были вынуждены мигрировать в область BI-приложений. Семейства продуктов СУБД, обеспечивающие возможности BI, действительно подталкивают рост рынка BI-платформ. Отчасти это происходит благодаря большей активности ряда поставщиков СУБД. Рассматривая различные инструменты, видим, что EBIS являются высоко функциональными средствами, но они не имеют такого большого значения, как BI-платформы или заказные BI-приложения. Зато BI-платформы обычно не так функционально полны, как корпоративные BI-наборы. При выборе BI-платформ нужно учитывать следующие характеристики: модульность, распределенную архитектуру, поддержку стандартов XML, OLE DB for OLAP, LDAP, CORBA, COM/DCOM и обеспечение работы в Web. Они должны также обеспечивать функциональность, специфическую для бизнес-интеллекта, а именно: доступ к БД (SQL), манипулирование многомерными данными, функции моделирования, статистический анализ и деловую графику. Эту категорию продуктов представляют фирмы Microsoft, SAS Institute, ORACLE, SAP и другие.

BI-приложения. В приложения бизнес-интеллекта часто встроены BI-инструменты (OLAP, генераторы запросов и отчетов, средства моделирования, статистического анализа, визуализации и data mining). Многие BI-приложения извлекают данные из ERP-приложений. BI-приложения обычно ориентированы на конкретную функцию организации или задачу, такие как анализ и прогноз продаж, финансовое бюджетирование, прогнозирование, анализ рисков, анализ тенденций, «churn analysis» в телекоммуникациях и т.п. Они могут применяться и более широко как в случае приложений управления эффективностью предприятия (enterprise perfomance management) или системы сбалансированных показателей (balanced scorecard).

Разведка данных (data mining) представляет собой процесс обнаружения корреляции, тенденций, шаблонов, связей и категорий . Она выполняется путем тщательного исследования данных с использованием технологий распознавания шаблонов, а также статистических и математических методов. При разведке данных многократно выполняются различные операции и преобразования над сырыми данными (отбор признаков, стратификация, кластеризация, визуализация и регрессия), которые предназначены: 1) для нахождения представлений, которые являются интуитивно понятными для людей, которые, в свою очередь, лучше понимают бизнес-процессы, лежащие в основе их деятельности; 2) для нахождения моделей, которые могут предсказать результат или значение определенных ситуаций, используя исторические или субъективные данные.

В отличие от использования OLAP разведка данных в значительно меньшей степени направляется пользователем, вместо этого полагается на специализированные алгоритмы, которые устанавливают соотношение информации и помогают распознать важные (и ранее неизвестные) тенденции, свободные от предвзятости и предположений пользователя.

Кроме перечисленных инструментов, в состав BI могут входить следующие средства анализа : пакеты статистического анализа и анализ временных рядов и оценки рисков; средства моделирования; пакеты для нейронных сетей; средства нечеткой логики и экспертные системы.

Дополнительно нужно отметить средства для графического оформления результатов : средства деловой и научно-технической графики; «приборные доски», средства аналитической картографии и топологических карт; средства визуализации многомерных данных.

Метаданные. Большинство BI-инструментов, представленных на рынке, используют слой метаданных или репозиторий. Бизнес-метаданные включают определения данных, которые хранятся в источниках данных, в терминах предметной области. Они также могут содержать правила и вычисления, которые должны быть определены для этого бизнеса. Кроме того, существуют технические метаданные для доступа к физическим данным. CASE-средства, реляционные СУБД, средства извлечения, преобразования и загрузки данных используют метаданные. При создании хранилища и витрин данных часто можно автоматически извлечь метаданные из источников данных, но иногда пользователям самим приходится доставать метаданные. Так, возможна сложная ситуация с несколькими репозиториями, существующими в одной организации. Отсутствие общих метаданных для инструментов - из-за отсутствия стандартов для метаданных - серьезная проблема для подразделений ИТ.

Плюсы и минусы технологии. Возможности пользователя по ведению многоаспектного оперативного анализа информации в терминах предметной области для поддержки принятия бизнес решений быстро расширяются. Параллельное движение от информационной анархии или диктатуры к информационной демократии расширяет контингент пользователей business intelligence. На первое место выходит потребность гибкого доступа к корпоративным данным, а не просто потребность решить конкретную функциональную задачу. Снижается прямая зависимость от подразделений ИТ, изготавливающих по заказу отчеты или запросы. Возможен переход от статических регламентных отчетов к «живому отчету», а наиболее продвинутые аналитики получают возможность проводить кросс-тематический анализ и построение сводных отчетов с нуля, имея семантических слой, описывающий все показатели и разрезы корпоративной информации. Эти же средства могут использовать программисты для быстрого создания регламентных, параметрических отчетов. Web-доступ к BI (как к статическому, так и к динамическому контенту) позволит обеспечить реальное корпоративное информационное пространство и коллективную работу сотрудников.

Основным риском является слишком быстрые изменения в технологии BI, использование непроверенных решений и средств. Нужно отслеживать поставщиков, оценивать их устойчивость, направления развития, регулярно пробовать новые средства, проводить типизацию и унификацию BI. Другой

риск связан с качеством данных - если они должным образом не преобразованы, не очищены и не консолидированы, то никакие «навороченные» возможности BI-инструментов или приложений не смогут увеличить достоверность данных. Ряд проблем могут возникнуть из-за не согласованности метаданных. В рамках большой корпорации эти вопросы решаются на инфраструктурном уровне путем создания корпоративного хранилища данных и централизованного управления метаданными. Создание хранилища поможет навести порядок в номенклатуре собираемых показателей, сборе данных, их распространении и санкционировании доступа. Сама BI-технология не в состоянии решить комплексно эти проблемы, а пренебрежение ими возвращает к информационной анархии и «силосным ямам данных» .

РАЗДЕЛ 3. ОСНОВНЫЕ ПОДХОДЫ И МЕТОДЫ ОПИСАНИЯ ИНФОРМАЦИОННЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ

Информация и данные. Виды и свойства информации. Сигналы и знаки. Классификация сигналов. Математические модели сигналов. Теория сигналов, семиотика, теория информации. Основные направления семиотики.

Основные понятия: информация, сообщение, сигнал

С большим основанием наше время можно назвать «веком информации». Все возрастающий поток информации обрушивается на человека. По существу вся деятельность человека неразрывно связана с получением и накоплением новой информации обокружающей среде. В эту среду входят естественные объекты живой и неживой природы и объекты, созданные человеком. Информация необходима человеку для удовлетворения своих все возрастающих потребностей.

Информация отражает одну из сторон реального мира и вместе с энергетическими и вещественными ресурсами образует триаду, необходимую для общественного развития. Если энергетические и вещественные потоки, образно говоря, питают некоторую систему, то потоки информации организуют ее функционирование, управляют ею.

Понятие информации, однако, в отличие от физических категорий массы и энергии, не приобрело однозначного толкования. Имеется большое количество определений понятия «информация», от наиболее простого, имеющего в своей основе конкретную практическую деятельность человека, до самого сложного, характеризующего информацию как философскую категорию.

Будем понимать под информацией (от лат. informatio – разъяснение, осведомление) совокупность каких-либо сведений, содержащих знания об изучаемом процессе или явлении. Эти сведения представляют определенный интерес для изучающего данный процесс или явление и поэтому становятся объектом хранения, передачи и преобразования. Будем полагать также, что имеется некоторый источник информации , обладающий способностью изменять во времени или пространстве свое состояние.

Информацию передают в виде сообщений. С оо б щен и е – это сведение о состоянии источника информации, выраженное в определенной форме и

предназначенное для передачи от источника информации к адресату. Отправителями и получателями сообщений могут быть как люди, так и технические устройства, которые регистрируют, хранят, преобразуют, передают и принимают информацию.

Для передачи сообщений используются сигналы. Сигналом (от лат. signum – знак) называется физический процесс, однозначно отображающий передаваемое сообщение о состоянии какого-либо объекта наблюдения и пригодный для передачи и обработки. Другими словами, под сигналом понимают материальный носитель сообщения.

Физическая природа сигнала может быть различная: звуковая, механическая, оптическая, электрическая или еще какая-либо.

Сообщения и соответствующие им сигналы могут быть дискретными или непрерывными. Дискретное сообщение представляет конечную последовательность отдельных символов. Дискретные сообщения характерны для телеграфии, передачи данных и телеметрии. Последовательность элементов дискретного сообщения преобразуется, например, в последовательность двоичных чисел. Для их передачи по каналу связи необходим сигнал с двумя признаками: 0 и 1. Роль таких сигналов могут выполнять посылки постоянного тока разной полярности, колебания с различными частотами и др.

Непрерывные сообщения представляют собой непрерывные процессы. Примерами непрерывных сообщений служат речь, музыка, телевизионное изображение и др.

С помощью специальных устройств непрерывные сообщения преобразуются в электрические непрерывные сигналы. Например, если сообщением является речь, то микрофон преобразует звуковые колебания воздушной среды в электрические колебания, которые и передаются по назначению.

Однако не всякий физический процесс следует считать сигналом. Сигналом является только тот физический процесс, который содержит информацию, то есть некоторую совокупность сведений о состоянии или положении некоторого объекта. В качестве сигнала, таким образом, можно рассматривать любой физический процесс при условии, что некоторые его параметры изменяются в соответствии с переносимым сообщением.

Сигналы можно разделить на естественные и специально создаваемые. К естественным сигналам относят, например, световые сигналы от физических объектов окружающего мира, космические сигналы, электрические поля биологических объектов и др. Примерами специально создаваемых могут служить сигналы, посылаемые радиолокационной станцией, сигналы лазеров для зондирования атмосферы и др.

По количеству физических переменных, характеризующих состояние источника информации, сигналы делят на одномерные , двумерные и трехмерные . Примерами одномерных сигналов являются ток в цепи микрофона, напряжение на выходе датчика температуры и др. Двумерные сигналы используются в случае, когда состояние источника определяется двумя переменными (координатами) одновременно. Двумерные сигналы, например, используются при обработке изображений для того, чтобы задать координаты точки на плоскости. При помощи трехмерных сигналов можно определить положение пространственных объектов или описать цветные изображения.

Передача информации заключается в переносе ее на расстояние при помощи сигналов различной физической природы.

Для того чтобы сделать сигнал объектом теоретического изучения или практических расчетов, необходимо указать способ его математического описания или, другими словами, создать математическую модель сигнала. Математической моделью называют совокупность математических соотношений, описывающих изучаемый процесс или явление. Математические модели позволяют анализировать свойства сигналов и синтезировать сигналы с требуемыми свойствами.

В простейшем случае математическая модель сигнала устанавливает соответствие между любым моментом времени t и величиной сигнала x, где

T – ограниченный или бесконечный интервал времени, называемый областью определения сигнала,

X – множество возможных значений сигнала. Это соответствие может быть задано в форме скалярной функции $x\ t\ (\ )$. В более сложных случаях модель содержит математические соотношения, которые характеризуют некоторые обобщенные свойства сигнала.

Создание математической модели – это первый и очень важный этап изучения физического процесса или явления. Во-первых, математическая модель позволяет абстрагироваться от конкретной физической природы носителя сигнала. При этом она приобретает определенную универсальность, то есть способность описывать различные по своей физической природе процессы или по техническому назначению объекты. Одна и та же математическая модель может описывать изменение тока, напряжения, давления, температуры и т.д.

Во-вторых, математическая модель создается так, чтобы она описывала именно те свойства сигнала, которые наиболее важны для конкретного исследования. Необходимо учитывать, что математическая модель может хорошо работать в одних условиях и быть совершенно неприемлемой в других. Любая математическая модель имеет свою область применения и эта область, как правило, может быть определена только в результате многократного применения модели. Поэтому при составлении математической модели большое значение приобретают экспериментальные исследования и практический опыт.

Большое значение при выборе математической модели сигнала имеет ее сложность. Математическая модель, с одной стороны, должна быть достаточно сложна, чтобы отображать существенные свойства изучаемого процесса или явления, а, с другой стороны, достаточно проста, чтобы использовать более простые математические методы анализа.

Очевидно, чем полнее математическая модель отражает свойства сигнала, тем шире область ее применения, тем больший круг задач позволяет она решать. В то же самое время, чем полнее математическая модель, тем она сложнее, тем больше трудностей следует ожидать при исследовании из-за необходимости привлекать более сложные математические методы. Поэтому математическая модель сигнала не должна содержать больше подробностей, чем это необходимо для решения данной задачи.

РАЗДЕЛ 4. ОСНОВЫ КЛАССИЧЕСКОЙ ТЕОРИИ ИНФОРМАЦИИ

Количество информации при конечном числе равновозможных исходов. Мера Хартли. Количество информации как случайная величина. Энтропия. Основные свойства энтропии. Среднее количество взаимной информации (дискретный случай). Энтропия объектов с непрерывным множеством состояний. Среднее количество взаимной информации (непрерывный случай).

Информационные характеристики источников сообщений. Источники дискретных сообщений. Энтропия источника дискретных сообщений. Понятие избыточности источника сообщений. Скорость создания информации источником дискретных сообщений. Источники непрерывных сообщений. Информационные характеристики источников непрерывных сообщений.

Информационные характеристики каналов связи. Понятие канала связи. Понятие скорости передачи и пропускной способности канала.

Общие сведения о каналах связи

Канал связи (англ. Channel, data line) – система технических средств и среда распространения сигналов для передачи сообщений (не только данных) от источника к получателю, и наоборот.

Канал связи, понимаемый в узком смысле, представляет только физическую среду распространения сигналов, например, физическую линию связи (рисунок 35).

От источника сообщения (говорящего человека) сообщение (речь) поступает на вход предающего устройства (микрофон). Передающее устройства преобразует сообщение в сигналы, которые поступают на вход канала связи.

На входе канала связи приемное устройство (телефонный капсюль) по принятому сигналу воспроизводит переданное сообщение, последнее воспринимается приемником сообщения (слушающим человеком).

Передатчик, канал связи, и приемник формируют систему передачи информации или систему связи .

По назначению системы связи разграничивают каналы телесигнализаций, телеизмерений, телеуправления (телекомандные), телеграфные, телефонные, звукового вещания, факсимильные, телевизионного вещания и т.д.

Каналы связи могут иметь много форм, включая каналы отвечающие требованиям хранения данных, которые могут передавать со-общения как только возникнет ситуация.

Примеры каналов связи включают:

- Соединения между инициирующим и оконечным узлами цели;
- Буфер на который сообщения могут быть положены и получены;
- Выделенный канал, обеспечиваемый средой либо физическим разделением, таким как многопарный кабель, либо электрическим разделением, таким как частотное уплотнение каналов связи или мультиплексирование с временным разделением каналов.
- Путь для перемещения электрического или электромагнитного сигнала, что обычно отличает от других параллельных путей.
- Часть записывающей среды, такой как дорожка или группа дорожек, что позволяет производить чтение или запись станции, или устройства звуковоспроизведения.
- В коммуникационных системах часть, что соединяет источников данных и переменных данный.

- Специфическая радиочастота, пара или диапазон частот, обычно обозначаемый буквой, номерам или кодовым словам и зачастую выделенная международным соглашением.
- пространство в Internet Relay Chat (IRC) сети, в которой участники могут связываться один с другим.

Все эти коммуникационные каналы разделяет то свойство, сто они переносят информацию. Информация переносится через канал сигналом.

Приведем несколько примеров:

Специфическая радио частота, пара или диапазон частот, обычно обозначаемый буквой, номером или кодовым словам и зачастую выделенная международным соглашением. Морское УКВ радио использует некие 88 каналов в УКВ диапазоне, для двунаправленной частотно-модулированной голосовой связи. Канал 16, для примера, означает частоту 156,800МГц.

В США еще семь каналов, WX1 – WX7, используются для оповещения погоды. Телевизионные каналы, такие как Североамериканский второй канал, расположен на частоте 55,25МГц, канал 13=211,25МГц. Каждый канал шириной 6МГц. Кроме этих физических каналов телевидение также имеет виртуальные каналы.

Wi-Fi состоит из нелицензированных каналов 1-13 в диапазоне от 2412МГц до 2484МГц с шагом в 5МГц.

Качественная характеристика

Сегодня, функционирование и развитие любого современного производства невозможно без использования автоматизированных систем управления (АСУ) и средств связи. В последнее десятилетие идет широкие внедрение вычислительной техники и средств передачи данных, обеспечивающих обмен информацией между вычислительными центрами объектами АСУ.

Существуют также распределенные вычислительные сети, построенные на основе выделенных аналоговых и цифровых каналов. Цифровые междугородные каналы еще слабо внедрены в существующую сеть, поэтому ее основу составляют аналоговые каналы.

В аналоговых сетях применяется такая же АПД, как и на ТфОП, за исключением того, что возможно использование 4-проводных модемов. В отличие от 19200 до 31200бит/с. Кроме того, использование выделенных каналов позволяет снизить затраты на междугородную передачу данных.

Для организаций, находящихся на небольшом расстоянии друг от друга (в переделах города), целесообразно использовать прямые провода. На сегодняшний день существует АПД, которая позволяет работать на скоростях от 64Кбит/с до 6Мбит/с на расстоянии нескольких километров. По стоимости она соизмерима с модемами для телефонных каналов, что делает этот вид связи очень перспективным.

Мобильная связь имеет ограниченное применение в распределенных вычислительных сетях (РВС), так как при этом возникают проблемы, связанные с распространением радиоволн: затухание сигнала, перерывы из-за переключения с одной частоты на другую при переходе из одной сотовой зоны в другую. Перечисленное выше, ограничивает скорость передачи данных до 4800бит/с.

Использование технологии ISDN дает абоненту возможность работать со скоростями от 64Кбит/с. Но пока еще этот вид связи имеет ограниченное распространение из-за высокой стоимости оборудования и трафика.

Самые высокие показатели качества и скорости передаваемой информации имеют волоконно-оптические линии связи (ВОЛС), так как в них используется отличная от модного кабеля среда распространения, что позволяет работать на них со скоростями в несколько сотен Гбит/с. При многих преимуществах ВОЛС имеет главный недостаток – высокую стоимость.

Если организация не имеет проложенного кабеля и стоимость его прокладки очень высока, возможно использование радиосвязи или спутниковой связи. Радиоканал позволяет передавать данные со скоростями от 64Кбит/с до 2,048Мьит/с.

Организация спутникового канала достаточно дорога. Прежде всего это стоимость оборудования (десятки тысяч долларов), а так-же повременная оплата канала связи.

Из вышесказанного следует, что модернизация существующих сетей связи требует значительных капиталовложений и не может носить революционный характер. Поэтому актуальность проблемы использования существующих средств связи, с учетом развивающихся информационных технологий, постоянно возрастает.

Поэтому построение глобальных корпоративных вычислительных сетей на основе существующих аналоговых каналов, делает проблему исследования качества передачи данных по таким каналам особенно актуальной.

Теоретическим и экспериментальным исследованиям среды передачи данных РВС посвящены многие труды, что подтверждает потребность в проведении дальнейших исследований, направленных на разработку способов оценки каналов тональной частоты (ТЧ), предназначенных для передачи данных в АСУ.

РАЗДЕЛ 5. ОСНОВЫ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Тема 5.1. ВОСПРИЯТИЕ ИНФОРМАЦИИ

Процесс восприятия информации и его особенности. Основные этапы восприятия информации. Первичное восприятие, обнаружение, распознавание, анализ информации. Схема процесса восприятия информации. Физический, морфологический, синтаксический и семантический аспекты восприятия.

Технология поиска информации. Поиск информации в Интернет. Поисковые машины, метапоисковые средства, онлайновые энциклопедии и справочники. Современные поисковые порталы. Извлечение ключевых слов из текстовых материалов в MS Word. Программы-экстракторы.

Системы автоматического анализа текста.

Задачи обработки текстов — неструктурированной документации, историй болезни, патентов и диссертаций и т. п. — можно разбить на две условные категории. К первой относятся задачи, с которыми ежедневно сталкивается любой пользователь: проверка орфографии, фильтрация спама, автоматический перевод небольших фрагментов текста (несколько предложений) и др. С точки зрения

исследователей в области автоматической обработки текстов (АОТ), все эти задачи почти решены, и сегодня более актуальны задачи из второй категории, требующие обработки больших текстовых массивов: нахождение релевантных ответов на вопросы (задачи «вопрос-ответ»), полноценный машинный перевод целостных текстов, анализ мнений и отзывов, конструирование рекомендательных систем, работающих с большими массивами неструктурированных данных. Отличительная особенность таких задач — их сложность и отсутствие формализации, приводящие к тому, что для них пока еще нет полноценного набора решений, а применяются вспомогательные методы выделения ключевых слов и словосочетаний, суммаризации (автоматического реферирования) текстов и классификации текстов. В этом ряду особое место занимают технологии визуализации больших объемов текстовых данных.

Теоретическую основу автоматической обработки текстов составляет компьютерная лингвистика, наиболее востребованы в которой методы машинного обучения, статистического анализа, модели Маркова, логические модели и модификации этих методов с учетом специфики Больших Данных [1]. Существует несколько подходов к такой модификации: распараллеливание алгоритмов, применение методов снижения размерности, предобработка данных, в ходе которой целостные тексты заменяются их отдельными элементами. Несмотря на различие между национальными языками, лингвистические методы могут быть универсальными — некоторые морфологические и синтаксические модели удается использовать для анализа текстов как на английском, так и на русском языке.

У компьютерной лингвистики и автоматической обработки текстов есть официальная дата рождения — январь 1954 года, когда был проведен Джорджтаунский эксперимент, в ходе которого машина перевела 60 простых предложений с русского на английский, что было вполне в духе времени. Следующий виток развития пришелся на 60-е годы — период эволюции идей искусственного интеллекта, и одной из ярких работ того времени стала программа Eliza, которая вела психотерапевтические беседы с пользователем, спрягая глаголы и манипулируя местоимениями. В 80-е годы компьютерная лингвистика пережила некоторое затишье, после которого начался бум корпусной лингвистики — исследований в области создания, поддержки и использования текстовых корпусов (множество разнообразных текстов, посвященных разным темам и написанных в различных жанрах и стилях). Появление компьютеров, на которых можно было целиком хранить, обрабатывать корпусы, содержащие наборы эталонных текстов, и проводить сложные вычисления, позволило активно использовать статистические методы и методы машинного обучения для работы с текстами. В целом в начале 90-х годов в области компьютерной лингвистики произошел переход к статистическим методам и, затем, методам машинного обучения и анализа данных, которые применяют к уже написанным и существующим текстам.

В настоящее время в области автоматической обработки текстов в России значительная часть работ посвящена переносу методов, разработанных для английского языка, на русский, и, к сожалению, оригинальных разработок очень мало.

Задачи компьютерной лингвистики

В сфере обработки текстов на сегодняшний день сформировалось два подхода: на основе моделей языка и правил, составленных экспертами; на базе

машинного обучения. Первый позволяет достичь лучших результатов, однако составление моделей и правил настолько трудоемкий процесс, что уступающие по качеству методы машинного обучения практически его вытеснили. Повышение качества достигается не за счет совершенствования математических методов, а за счет увеличения и улучшения обучающей выборки. Оба подхода направлены сегодня на решение следующих задач.

Анализ и градация мнений. Соотнесение текста, написанного от первого лица, с дискретной шкалой оценок: плохо, хорошо, очень хорошо и т. д. Используется для анализа отзывов в интернет-магазинах и высказываний в социальных сетях.

Анализ тональности высказываний. Выявление позитивного или негативного отношения к обсуждаемому предмету. Используется для анализа отзывов, генерации диалога и т. д.

Классификация текстов по темам. Отнесение текста к той или иной тематике. Используется во многих приложениях — в частности, в рекомендательных системах, для рубрикации текстов в онлайн-библиотеках и для организации новостных потоков.

Генерация речи. Используется в робототехнике, смартфонах, навигаторах.

Ведение диалога. Анализ реплик собеседника и формирование на их основе ответов. Используется в робототехнике, экспертных системах — например, Королевский банк Шотландии частично заменил контакт-центры роботами, поддерживающими диалог с пользователем.

Проверка правописания. Используется в текстовых редакторах, поисковых системах.

Извлечение смысла из текста. Выделение ключевых слов и словосочетаний, трендов, суммаризация. Применяется в новостных системах для агрегирования серии новостных сообщений, базах знаний для организации хранения знаний и вывода новых фактов.

Поиск ответов на вопросы. Подборка по вопросу и, возможно, контексту наиболее релевантного ответа. Применяется в поисковых и экспертных системах.

Машинный перевод.

Системы АОТ можно классифицировать по виду лицензирования (проприетарные — как правило, принадлежащие известным производителям, и академические разработки — распространяемые бесплатно); открытости (системы могут быть либо доступны только узкому кругу людей, либо находиться в открытом доступе); целевой аудитории (исследователи в области компьютерной лингвистики, разработчики, рядовые пользователи и т. п., что определяет интерфейс системы); мультиязычности (различие по числу поддерживаемых языков); характеру (готовые системы или библиотеки инструментов обработки текстов); универсализму (решение конкретных задач, обработка текстов в целом); используемым данным (тип и объемы обрабатываемых данных); применяемым экспертным правилам и математическим моделям; ориентации на конкретную прикладную область.

Мультиязычные системы часто более коммерчески привлекательны и просты в использовании. В свою очередь, системы, ориентированные на конкретный язык или подмножество языков, обеспечивают пусть и небольшой, но

очень важный во многих задачах прирост качества за счет учета специфики языка. Классический пример мультиязычной системы — переводчик Google.

Системы, рассчитанные на достаточно широкое (и, как правило, коммерческое) использование, обладают хорошо развитым интерфейсом для конечного пользователя (например, Microsoft Bing Translator и Google translator, ОРФО, программа для автоматического переключения между различными раскладками клавиатуры Punto Switcher, различные утилиты «Яндекса» и т. д.). Ряд этих систем обладает также своим собственным программным интерфейсом (ОРФО, Microsoft Bing Translator). Но в данном случае он является скорее приятным дополнением, чем основным способом использования систем. Напротив, для систем, рассчитанных только на исследователей или являющихся составной частью более объемных проектов, программный интерфейс становится главным (а часто и единственным) способом взаимодействия. Интерфейсы для конечных пользователей в этих системах рассчитаны скорее на работу в тестовом режиме и часто являются консольными. В качестве примеров такого рода систем можно назвать mystem, АОТ, pyMorphy 1 и 2, «Томита парсер» [4], OpenXerox, Snowball. Почти все они предназначены для решения конкретных задач, возникающих на различных этапах анализа текстов: выделения слов из текста (токенизация), морфологического анализа (определения частей речи и других грамматических характеристик), построения синтаксической структуры предложений и т. д.

Корпусы — неотъемлемая часть многих систем обработки текстов. Каждое слово в корпусах снабжено исчерпывающими грамматическими характеристиками: к какой части речи оно принадлежит, в какой форме оно находится, какова его синтаксическая роль. Корпусы служат входными данными для обучения в задачах классификации текстов по темам и жанрам, для обучения синтаксических парсеров и программ, используемых для снятия омонимии и разрешения анафоры. Параллельные корпусы, состоящие из одинаковых текстов на разных языках, используют для обучения машинных переводчиков. Как правило, корпусы собираются десятилетиями, и в их создании участвуют большие исследовательские группы — например, проект «Национальный корпус русского языка» существует уже 13 лет и поддерживается компанией «Яндекс».

Важный тип входных данных любой системы АОТ — морфологические словари. Например, библиотека «АОТ», используемая во многих исследовательских и коммерческих проектах, представляет собой словарь Зализняка в цифровой форме. Тезаурусы (или семантические сети) — другой тип широко востребованных входных данных. Пожалуй, самый известный тезаурус — это WordNet, представляющий собой ресурс, в котором слова связаны с помощью так называемых семантических отношений: синонимии, гиперонимии (частное — обобщение), гипонимии (обобщение — частное), меронимии (часть — целое) и др. WordNet полезен в задачах машинного перевода, генерации текстов, классификации текстов. К сожалению, русского аналога WordNet пока нет.

Решение практически любой задачи АОТ так или иначе включает в себя проведение анализа текста на нескольких уровнях представления.

Графематический анализ. Выделение из массива данных предложений и слов (токенов).

Морфологический анализ. Выделение грамматической основы слова, определение частей речи, приведение слова к словарной форме.

Синтаксический анализ. Выявление синтаксических связей между словами в предложении, построение синтаксической структуры предложения.

Семантический анализ.] Выявление семантических связей между словами и синтаксическими группами, извлечение семантических отношений.

Каждый такой анализ — самостоятельная задача, не имеющая собственного практического применения, но активно используемая для решения более общих задач. Многие исследовательские системы предназначены для решения именно вспомогательных задач. Такие системы применяются либо для апробации методов и проведения вычислительных экспериментов, либо в качестве составных частей (или библиотек) для систем, решающих ту или иную прикладную задачу. Примером таких систем могут служить средство NLTK для графематического анализа и токенизации, морфологический анализатор mystem и синтаксический парсер ЭТАП3.

Универсализм в АОТ подразумевает наличие в системе набора взаимосвязанных методов и подходов. Существует два класса таких систем. К первому относятся системы, разрабатываемые исследовательскими департаментами крупных компаний: IBM, Intel, SAS, ABBYY, Microsoft, Xerox и т. д. В качестве примеров систем, предназначенных для обработки текстов на английском языке, можно назвать IBM Content Analytics, SAS Text Miner и IBM Watson. Ко второму классу относятся открытые интегрированные программные пакеты, созданные в университетах и представляющие собой множество методов и моделей, построенных на единой программной и математической платформе. Для английского языка можно назвать системы Apache OpenNLP, StanfordNLP, NLTK, GATE. Систем для работы с русским языком, претендующих на универсализм, пока нет, более того, в случае русского языка отсутствуют даже доступные для конечного пользователя системы, решающие основные лингвистические задачи: выделение ключевых слов, классификация текстов по темам, определение тональности текстов. В таблице перечислены программные системы, работающие с русским языком.

Некоторые системы АОТ направлены на анализ текстов определенных жанров или тематики. Например, система Watson применяется в медицине для диагностирования и облегчения процедуры принятия врачами решений. Рекомендательная система новостных сообщений News360 представляет собой приложение для мобильных устройств, с помощью которого пользователь может читать и выбирать наиболее интересную для него информацию. На основе предпочтений пользователя система предлагает новые статьи, собранные с разных новостных порталов и отвечающие конкретной тематике. В некоторых случаях эти системы умеют определять тональность новостного сообщения — например, пользователь может просматривать только хорошие новости и исключить из своей ленты все плохие. Рекомендательные системы, работающие с текстовыми данными, особенно востребованы в интернет-магазинах. С точки зрения АОТ отзыв пользователя интернет-магазина — это текст, имеющий явную тональную окраску и посвященный конкретному предмету. По отзыву пользователя необходимо определить, остался ли он доволен купленным товаром или нет, а если ему что-то не понравилось, то понять, что именно. Кроме того, перед интернет-магазинами встает задача выявления поддельных отзывов, написанных производителем товара. Создателям специализированных систем анализа отзывов

приходится идти на компромисс — если специализация системы слишком узкая (например, она нацелена только на тексты про мультиварки), то ее невозможно будет использовать для анализа текстов другой специализации (косметика).

Сегодня многие модели, разработанные в недрах научных сообществ, взяты на вооружение крупными игроками рынка ИТ (Google, IBM, Microsoft), однако в секторе, ориентированном на работу с русским языком, наблюдается ощутимое отставание от английского, китайского, арабского и от европейских языков. Существующие системы решают либо совсем простые (проверка орфографии, базовая корректировка поискового запроса), либо вспомогательные (выделение основы слов, приведение слова к начальной форме), либо специальные задачи (автоматическое составление резюме, анализ компетенций, анализ профиля среднестатистического пользователя социальной сети). Сравнение с рядом славянских и восточно-европейских языков также оказывается не в пользу русского — в последние годы очень активно ведутся исследования (включающие в себя составление корпусов и разработку оригинальных методов) в области обработки текстов на чешском, болгарском, румынском, польском языках.

Одни из лучших считаются следующие системы анализа текстов:

Megaputer PolyAnalyst от Мегапьютер Интеллидженс

PolyAnalyst – это программная платформа визуальной разработки сценариев анализа данных и текстов, а также построения интерактивных отчётов, не требующая навыков программирования для аналитики. Программный продукт PolyAnalyst (рус. Полианалист) от компании Megaputer предназначена для анализа структурированных и неструктурированных данных на высокопрофессиональном промышленном уровне. Система содержит более 100 функциональных узлов, в которых реализованы алгоритмы интеллектуального анализа данных и обработки естественного языка (NLP), а также средства загрузки данных в более чем 50 форматах, визуализации результатов, построения и публикации отчётов. Аналитическая программа может быть без обучения и без наличия профессиональных навыков анализа больших данных использована отраслевыми специалистами, в том числе юристами, финансистами, маркетологами, бизнес-аналитиками.

Программное обеспечение PolyAnalyst обладает обширными инструментами работы с неструктурированными текстовыми данными. PolyAnalyst позволяет проводить классификацию и кластеризацию текстов, извлекать сущности и факты, определять тематику и тональность, а также решать ряд других прикладных аналитических задач по работе с документами на русском, английском и других языках.

Аналитическая платформа PolyAnalyst предлагает простой и интуитивно-понятный пользовательский интерфейс. Аналитические сценарии строятся путём перетаскивания элементов и их соединения друг с другом на поле разработки решения. Элементы сценария представляют собой действия по загрузке, манипуляции и анализу данных. Пользователь формирует наглядную, легко модифицируемую и расширяемую структуру скрипта сценария автоматического анализа. Построенные аналитические сценарии могут быть динамически привязаны к ключевым бизнес-процессам компании, в ходе которых проводится загрузка, предобработка, анализ и визуализация данных. Сценарии анализа также

могут интегрироваться в текущие ИТ-системы компании путём обмена данными через программный интерфейс (API).

Платформа может быть установлена как в исполнении настольной программы для ПК, так и в клиент-серверной многопользовательской конфигурации в закрытый защищённый контур компании. Клиентский доступ в таком случае осуществляется или через запуск приложения операционной системы, или через тонкий клиент в веб-браузере.

M-Brain Intelligence Plaza – это ИТ-платформа для управления потоками информации о рынках и конкурентах для отделов аналитики, продаж, маркетинга, менеджмента. Хранение в облаке, структурирование и внутрикорпоративная рассылка информации по темам.

Инлексис Голосовой бот – это интеллектуальный сервис для эффективного обзвона клиентов, позволяющий крупному бизнесу экономить миллионы рублей на сокращении операционных расходов.

IQPLATFORM – это цифровая аналитическая платформа, позволяет выполнять продвинутую аналитику на базе больших объёмов информации, синтез новых знаний и мониторинг и контроль информационных объектов.

Elasticsearch от Elastic NV

Платформа Elasticsearch – это программное обеспечение с открытым исходным кодом, предназначенное для поиска, сбора, анализа и хранения текстовых данных с использованием интеллектуальных алгоритмов.

Тема 5.2. ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ

Цели и виды преобразования информации. Редукция, кодирование, модуляция. Дискретизация сигнала во времени. Основные методы дискретизации сигнала. Оценка погрешности дискретизации. Квантование сигнала по уровню. Дисперсия шума квантования. Цифровое представление информации. Двоичная, восьмеричная и шестнадцатеричная формы представления.

Кодирование информации. Статистическое и помехоустойчивое кодирование. Шифрование данных. Основные криптографические методы. Симметричные алгоритмы шифрования. Алгоритмы шифрования с открытым ключом. Алгоритм RSA. Алгоритм Эль-Гамаля.

Шифрование и расшифрование

Предположим, что отправитель хочет послать сообщение получателю. Более того, отправитель желает засекретить это сообщение, чтобы никто, кроме получателя, не смог его прочитать.

Сообщение состоит из открытого текста. Процесс преобразования открытого текста с целью сделать непонятным его смысл для посторонних называется шифрованием. В результате шифрования сообщения получается шифртекст. Процесс обратного преобразования шифртекста в открытый текст называется расшифрованием.

Наука, которая учит, как следует поступать, чтобы сохранить содержание сообщений в тайне, называется криптографией. Людей, занимающихся криптографией, зовут криптографами. Криптоаналитики являются специалистами в области криптоанализа — науки о вскрытии шифров, которая отвечает на вопрос о

том, как прочесть открытый текст, скрывающийся под шифрованным. Раздел науки, объединяющий криптографию и криптоанализ, именуется криптологией.

Обозначим открытый текст буквой P (от английского слова plaintext). Это может быть текстовый файл, битовое изображение, оцифрованный звук - что угодно. Единственное ограничение связано с тем, что, поскольку предметом изложения является компьютерная криптография, под P понимаются исключительно двоичные данные.

Шифртекст обозначается буквой C (от английского слова ciphertext) и также представляет собой двоичные данные. Объем полученного шифртекста иногда совпадает с объемом соответствующего открытого текста, а иногда превышает его. После зашифрования преобразованный открытый текст может быть передан по каналам компьютерной сети или сохранен в памяти компьютера.

На вход функции шифрования E подается P, чтобы на выходе получить C. В обозначениях, принятых в математике, это записывается как:

E(P) = C

При обратном преобразовании шифртекста в открытый текст на вход функции расшифрования D поступает C, а на выходе получается P:

D(C) = P

Поскольку смысл любого криптографического преобразования открытого текста состоит в том, чтобы потом этот открытый текст можно было восстановить в первозданном виде, верно следующее соотношение:

D(E(P)) = P

Шифры и ключи

Криптографический алгоритм, также называемый шифром или алгоритмом шифрования, представляет собой математическую функцию, используемую для шифрования и расшифрования. Если быть более точным, таких функций две: одна применяется для шифрования, а другая — для расшифрования.

Когда надежность криптографического алгоритма обеспечивается за счет сохранения в тайне сути самого алгоритма, такой алгоритм шифрования называется ограниченным. Ограниченные алгоритмы представляют значительный интерес с точки зрения истории криптографии, однако совершенно непригодны при современных требованиях, предъявляемых к шифрованию. Ведь в этом случае каждая группа пользователей, желающих обмениваться секретными сообщениями, должна обзавестись своим оригинальным алгоритмом шифрования. Применение готового оборудования и стандартных программ исключено, поскольку тогда любой сможет приобрести это оборудование и эти программы и ознакомиться с заложенным в них алгоритмом шифрования. Придется разрабатывать собственный криптографический алгоритм, причем делать это надо будет каждый раз, когда кто-то из пользователей группы захочет ее покинуть или когда детали алгоритма случайно станут известны посторонним.

В современной криптографии эти проблемы решаются с помощью использования ключа, который обозначается буквой K (от английского слова key). Ключ должен выбираться среди значений, принадлежащих множеству, которое называется ключевым пространством. И функция шифрования E, и функция расшифрования D зависят от ключа. Сей факт выражается присутствием K в качестве подстрочного индекса у функций E и D:

$E_K(P)=C$

DK(C)=P

По-прежнему справедливо следующее тождество:

DK(EK(P))=P

Некоторые алгоритмы шифрования используют различные ключи для шифрования и расшифрования. Это означает, что ключ шифрования К1 отличается от ключа расшифрования К2. В этом случае справедливы следующие соотношения:

EK1(P)=C

DK2(C)=P

DK(EK(P))=P

Надежность алгоритма шифрования с использованием ключей достигается за счет их надлежащего выбора и последующего хранения в строжайшем секрете. Это означает, что такой алгоритм не требуется держать в тайне. Можно организовать массовое производство криптографических средств, в основу функционирования которых положен данный алгоритм. Знание криптографического алгоритма не позволит злоумышленнику прочесть зашифрованные сообщения, поскольку он не знает секретный ключ, использованный для их зашифрования.

Под криптосистемой понимается алгоритм шифрования, а также множество всевозможных ключей, открытых и шифрованных текстов.

Симметричные алгоритмы шифрования

Существуют две разновидности алгоритмов шифрования с использованием ключей — симметричные и с открытым ключом. Симметричным называют криптографический алгоритм, в котором ключ, используемый для шифрования сообщений, может быть получен из ключа расшифрования и наоборот. В большинстве симметричных алгоритмов применяют всего один ключ. Такие алгоритмы именуются одноключевыми, или алгоритмами с секретным ключом, и требуют, чтобы отправитель сообщений и их получатель заранее условились о том, каким ключом они будут пользоваться. Надежность одно-ключевого алгоритма определяется выбором ключа, поскольку его знание дает возможность злоумышленнику без помех расшифровывать все перехваченные сообщения. Поэтому выбранный ключ следует хранить в тайне от посторонних.

Шифрование и расшифрование в симметричных криптографических алгоритмах задаются уже знакомыми формулами:

EK(P)=C

DK(C)=P

Симметричные алгоритмы шифрования бывают двух видов. Одни из них обрабатывают открытый текст побитно. Они называются потоковыми алгоритмами, или потоковыми шифрами. Согласно другим, открытый текст разбивается на блоки, состоящие из нескольких бит. Такие алгоритмы называются блочными, или блочными шифрами. В современных компьютерных алгоритмах блочного шифрования обычно длина блока составляет 64 бита.

Алгоритмы шифрования с открытым ключом

Алгоритмы шифрования с открытым ключом, также называемые асимметричными алгоритмами шифрования, устроены так, что ключ, используемый для шифрования сообщений, отличается от ключа, применяемого для их расшифрования. Более того, ключ расшифрования не может быть за обозримое время вычислен, исходя из ключа шифрования. Свое название

алгоритмы с открытым ключом получили благодаря тому, что ключ шифрования не требуется держать в тайне. Любой может им воспользоваться, чтобы зашифровать свое сообщение, но только обладатель соответствующего секретного ключа расшифрования будет в состоянии прочесть это шифрованное сообщение. Ключ шифрования обычно называют открытым ключом, а ключ расшифрования — тайным ключом. Иногда тайный ключ называют также j секретным, однако чтобы избежать путаницы с симметричными алгоритмами, это название не будет использоваться при дальнейшем изложении.

Несмотря на тот факт, что сообщения шифруются с помощью открытого ключа, а расшифровываются с помощью тайного ключа, процесс шифрования и расшифрования все равно записывается так:

EK(P)=C

DK(C)=P

Иногда сообщения шифруются с использованием тайного ключа, а расшифровываются посредством открытого ключа. Несмотря на возможную путаницу, этот факт математически по-прежнему выражается в виде:

EK(P)=C

DK(C)=P

Шифры замены

Шифром замены называется алгоритм шифрования, который производит замену каждой буквы открытого текста на какой-то символ шифрованного текста. Получатель сообщения расшифровывает его путем обратной замены.

В классической криптографии различают 4 разновидности шифров замены:

- Простая замена, или одноалфавитный шифр. Каждая буква открытого текста заменяется на один и тот же символ шифртекста.
- Омофонная замена. Аналогична простой замене с единственным отличием: каждой букве открытого текста ставятся в соответствие несколько символов шифртекста. Например, буква "А" заменяется на цифру 5, 13, 25 или 57, а буква "Б" — на 7, 19, 31 или 43 и так далее.
- Блочная замена. Шифрование открытого текста производится блоками. Например, блоку "АБА" может соответствовать "РТК", а блоку "АББ" — "СИЛ".
- Многоалфавитная замена. Состоит из нескольких шифров простой замены. Например, могут использоваться пять шифров простой замены, а какой из них конкретно применяется для шифрования данной буквы открытого текста, — зависит от ее положения в тексте.

Примером шифра простой замены может служить программа ROT13, которую обычно можно-найти в операционной системе UNIX. С ее помощью буква "A" открытого текста на английском языке заменяется на букву "N", "B" — на "O" и так далее. Таким образом, ROT13 циклически сдвигает каждую букву английского алфавита на 13 позиций вправо. Чтобы получить исходный открытый текст надо применить функцию шифрования ROT13 дважды:

P = ROT13(ROT13(P))

Все упомянутые шифры замены легко взламываются с использованием современных компьютеров, поскольку замена недостаточно хорошо маскирует стандартные частоты встречаемости букв в открытом тексте.

Разновидностью шифра замены можно считать код, который вместо букв осуществляет замену слов, фраз и даже целых предложений. Например, кодовый текст "ЛЕДЕНЕЦ" может соответствовать фразе открытого текста "ПОВЕРНУТЬ ВПРАВО НА 90°". Однако коды применимы только при определенных условиях: если, например, в коде отсутствует соответствующее значение для слова "МУРАВЬЕД", то вы не можете использовать это слово в открытом тексте своего сообщения, предназначенном для кодирования.

Шифры перестановки

В шифре перестановки буквы открытого текста не замещаются на другие, а меняется сам порядок их следования. Например, в шифре простой колонной перестановки исходный открытый текст записывается построчно (число букв в строке фиксировано), а шифртекст получается считыванием букв по колонкам. Расшифрование производится аналогично: шифртекст записывается поколонно, а открытый текст можно затем прочесть по горизонтали.

Для повышения стойкости полученный шифртекст можно подать на вход второго шифра перестановки. Существуют еще более сложные шифры перестановки, однако почти все они легко взламываются с помощью компьютера.

Хотя во многих современных криптографических алгоритмах и используется перестановка, ее применение ограничено узкими рамками, поскольку в этом случае требуется память большого объема, а также накладываются ограничения на длину шифруемых сообщений. Замена получила значительно большее распространение.

Операция сложения по модулю 2

Операция сложения по модулю 2, которая в языке программирования С обозначается знаком ^, а в математике — знаком + в кружочке, представляет собой стандартную операцию над битами:

С помощью сложения по модулю 2 можно выполнять многоалфавитную замену, прибавляя к битам ключа соответствующие биты открытого текста. Этот алгоритм шифрования является симметричным. Поскольку двойное прибавление одной и той же величины по модулю 2 восстанавливает исходное значение, шифрование и расшифрование выполняются одной и той же программой.

К сожалению, данный алгоритм обладает очень слабой стойкостью. Тем не менее АНБ одобрило его использование в цифровых сотовых телефонах американских производителей для засекречивания речевых переговоров. Он также часто встречается в различных коммерческих программных продуктах.

Следует помнить, что с помощью данного алгоритма вы можете надежно спрятать от человека, не сведущего в криптоанализе, содержание своей переписки, однако опытному криптоаналитику понадобится всего несколько минут, чтобы его взломать. Делается это очень просто, если предположить, например, что открытый текст сообщения написан на английском языке.

1. Сначала следует определить длину ключа. Шифртекст последовательно складывается по модулю 2 со своей копией, сдвинутой на различное число байт, и в полученном векторе подсчитывается число совпадающих компонентов. Когда величина сдвига кратна длине ключа, это число превысит 6% от общей длины исследуемого шифртекста. Если не кратна, то совпадений будет меньше 0,4%. Проанализировав полученные данные, можно сделать обоснованный вывод о длине ключа.

2. Затем надо сложить шифртекст по модулю 2 со своей копией, сдвинутой на величину длины ключа. Эта операция аннулирует ключ и оставит в наличии открытый текст сообщения, сложенный по модулю 2 со своей копией, сдвинутой на величину длины ключа.

Компьютерные алгоритмы шифрования

Существует великое множество алгоритмов шифрования, придуманных специально в расчете на реализацию в виде компьютерных программ. Среди наиболее известных можно упомянуть:

1) Data Encryption Standard (DES). Симметричный алгоритм шифрования, являющийся в США государственным стандартом.

2) RSA. Алгоритм шифрования с открытым ключом, названный по первым буквам фамилий его создателей (Rivest, Shamir, Adleman).

3) ГОСТ 28147-89. Симметричный алгоритм шифрования, одобренный сначала в СССР, а затем и в России для использования в качестве государственного стандарта.

Криптоаналитические атаки

Криптография ставит своей целью сохранение переписки в тайне от посторонних людей, которые захотят с ней ознакомиться. Таких людей криптографы называют злоумышленниками, противниками, перехватчиками или просто врагами. При этом предполагается, что они могут перехватывать любые сообщения, которыми обмениваются отправитель и получатель.

Криптоанализ заключается в получении доступа к открытому тексту шифрованного сообщения. В ходе успешного криптоаналитического исследования криптосистемы может быть найден не только открытый текст, но и сам ключ. Криптоаналитик занимается поисками слабостей в криптосистеме, которые могут позволить ему прочесть шифрованное сообщение, или отыскать ключ, или и то, и другое вместе. Если противник узнал ключ не с помощью криптоанализа, а каким-то другим способом (выкрал или купил), то говорят, что ключ был скомпрометирован.

Попытка криптоанализа называется атакой. Успешная криптоаналитическая атака зовется взломом, или вскрытием.

В современной криптологии принято считать, что надежность шифра определяется только секретностью используемого ключа. Правило, впервые сформулированное голландцем А. Керкхоффом (1835—1903), гласит о том, что весь механизм шифрования, за исключением значения ключа, предположительно известен противнику. Это предположение является довольно естественным. Например, хотя ФАПСИ вряд ли знакомит АНБ со своими криптографическими алгоритмами, какое-то представление о них там все равно имеется. Следовательно, правило Керкхоффа является относительно хорошим допущением при рассмотрении надежности алгоритмов шифрования. Если шифр невозможно взломать, зная абсолютно все детали алгоритма шифрования, значит это тем более нельзя сделать, не обладая подобными знаниями во всей их полноте.

Известны 4 основных типа криптоаналитических атак. При рассмотрении каждой из них подразумевается, что криптоаналитик в курсе всех деталей подвергаемого криптоанализу алгоритма шифрования.

1. Атака со знанием только шифртекста. В распоряжении криптоаналитика имеются несколько сообщений, которые были зашифрованы с

использованием одного и того же алгоритма шифрования. Задача криптоаналитика состоит в нахождении открытого текста наибольшего числа перехваченных сообщений. Он может также попытаться найти ключи, которые применялись для шифрования этих сообщений, чтобы потом прочесть другие сообщения, зашифрованные с использованием тех же ключей.

Дано: C1 = EK1(P1), C2 = EK2(P2) … Ci = EKi(Pi)

Найти: P1, P2, …, Pi или K1, K 2, …, K i

2. Атака со знанием открытого текста. Криптоаналитик имеет доступ не только к шифрованным текстам нескольких сообщений, но и знает их открытые тексты. От него требуется найти ключи, которые использовались для шифрования этих сообщений.

Дано: P1, C1 = EK1(P1), P2,C2 = EK2(P2) … Pi, Ci = EKi(Pi)

Найти: K1, K 2, …, K i

3. Атака с выбранным открытым текстом. Криптоаналитик не только знает шифрованные и открытые тексты нескольких сообщений, но и может определять содержание этих сообщений. Данная разновидность криптоаналитической атаки является более мощной по сравнению с атакой со знанием открытого текста, поскольку здесь криптоаналитик может по своему усмотрению выбирать открытый текст, подлежащий зашифрованию, и, тем самым, получать больше информации об используемых ключах. Его задача по-прежнему состоит в нахождении ключей.

Дано: P1, C1 = EK1(P1), P2,C2 = EK2(P2) … Pi, где Ci = EKi(Pi), где P1, P2,…

Pi – выбраны криптоаналитиком.

Найти: K1, K 2, …, K i

4. Адаптивная атака с выбранным открытым текстом. Эта атака является разновидностью атаки с выбранным открытым текстом. Криптоаналитик не только выбирает открытые тексты посылаемых шифрованных сообщений, но и может менять свой выбор в зависимости от результатов их шифрования.

Имеются по крайней мере еще 3 разновидности криптоаналитических атак.

1. Атака с выбранным шифртекстом. Криптоаналитику предоставлена возможность выбора шифртекстов, подлежащих расшифрованию получателем. Он также имеет доступ к соответствующим открытым текстам. Требуется найти ключи.

Этой криптоаналитической атаке, как правило, подвергаются алгоритмы шифрования с открытым ключом. Хотя иногда она эффективна и против симметричных криптосистем. Атаку с выбранным открытым текстом и с выбранным шифртектом называют атакой с выбранным текстом.

2. Атака с выбранным ключом. В ходе этой атаки Криптоаналитик обладает некоторыми знаниями относительно правил, по которым отправитель и получатель сообщений выбирают ключи шифрования.

3. Атака с применением грубой силы. Криптоаналитик занимается подкупом, шантажом или пытками, чтобы получить сведения, необходимые ему для взлома криптосистемы. Подкуп иногда выделяют в отдельную категорию и называют атакой с покупкой ключа. Эти атаки являются очень эффективными и зачастую предоставляют наиболее легкий путь для получения доступа к открытым текстам шифрованных сообщений.

Атаки со знанием открытого текста и с выбранным открытым текстом не так уж редко встречаются на практике, как можно подумать. Известны случаи, когда криптоаналитику удавалось подкупить шифровальщика, чтобы он зашифровал сообщение, открытый текст которого известен криптоаналитику. Иногда даже не требуется никого подкупать, поскольку открытые тексты многих сообщений начинаются и заканчиваются стандартными фразами.

С этой точки зрения зашифрованная программа на языке С особенно уязвима, поскольку содержит множество зарезервированных слов типа #oef; г •, #include, if, then И do.

Не следует забывать и о правиле Керкхоффа. Попытка добиться высоко:; надежности криптографического алгоритма за счет сохранения в rail не принципов его работы является малопродуктивной. Криптоаналитик может выполнить дизассемблирование любой сверхсложной программы шифрования и методом обратного проектирования воспроизвести алгоритм, положенный в основу ее функционирования. Такое случается довольно часто. Лучшие алгоритмы шифрования являются общественным достоянием уже в течение многих лет, и над их взломом продолжают безуспешно трудиться самые способные криптоаналитики в мире.

Шифрование в каналах связи компьютерной сети

Одной из отличительных характеристик любой компьютерной сети является ее деление на так называемые уровни, каждый из которых отвечает за соблюдение определенных условий и выполнение функций, необходимых для общения между компьютерами, связанными в сеть. Это деление на уровни имеет фундаментальное значение для создания стандартных компьютерных сетей. Поэтому в 1984 г. несколько международных организаций и комитетов объединили свои усилия и выработали примерную модель компьютерной сети, известную под названием OSI (Open Systems Interconnection — Модель открытых сетевых соединений).

Согласно модели OSI коммуникационные функции разнесены по уровням. Функции каждого уровня независимы от функций ниже- и вышележащих уровней. Каждый уровень может непосредственно общаться только с двумя соседними. Модель OSI определяет 7 уровней: верхние 3 служат для связи с конечным пользователем, а нижние 4 ориентированы на выполнение коммуникационных функций в реальном масштабе времени.

Теоретически шифрование данных для передачи по каналам связи компьютерной сети может осуществляться на любом уровне модели OSI. На практике это обычно делается либо на самых нижних, либо на самых верхних уровнях. Если данные шифруются на нижних уровнях, шифрование называется канальным, а если на верхних, то такое шифрование называется сквозным. Оба этих подхода к шифрованию данных имеют свои преимущества и недостатки.

Канальное шифрование

При канальном шифровании шифруются абсолютно все данные, проходящие по каждому каналу связи, включая открытый текст сообщения, а также информацию о его маршрутизации и об используемом коммуникационном протоколе. Однако в этом случае любой интеллектуальный сетевой узел (например, коммутатор) будет вынужден расшифровывать входящий поток данных, чтобы соответствующим образом его обработать, снова зашифровать и передать на другой узел сети.

Тем не менее канальное шифрование представляет собой очень эффективное средство защиты информации в компьютерных сетях. Поскольку шифрованию подлежат все данные, передаваемые от одного узла сети к другому, у криптоаналитика нет никакой дополнительной информации о том, кто служит источником этих данных, кому они предназначены, какова их структура и т. д. А если еще позаботиться и о том, чтобы, пока канал простаивает, передавать по нему случайную битовую последовательность, сторонний наблюдатель не сможет даже сказать, где начинается и где заканчивается текст передаваемого сообщения.

Не слишком сложной является и работа с ключами. Одинаковыми ключами следует снабдить только два соседних узла сети связи, которые затем могут менять используемые ключи независимо от других пар узлов.

Самый большой недостаток канального шифрования заключается в том, что данные приходится шифровать при передаче по каждому физическому каналу компьютерной сети. Отправка информации в незашифрованном виде по какому-то из каналов ставит под угрозу обеспечение безопасности всей сети. В результате стоимость реализации канального шифрования в больших сетях может оказаться чрезмерно высокой.

Кроме того, при использовании канального шифрования дополнительно потребуется защищать каждый узел компьютерной сети, по которому передаются данные. Если абоненты сети полностью доверяют друг другу и каждый ее узел размещен там, где он защищен от злоумышленников, на этот недостаток канального шифрования можно не обращать внимания. Однако на практике такое положение встречается чрезвычайно редко. Ведь в каждой фирме есть конфиденциальные данные, знакомиться с которыми могут только сотрудники одного определенного отдела, а за его пределами доступ к этим данным необходимо ограничивать до минимума.

Сквозное шифрование

При сквозном шифровании криптографический алгоритм реализуется на дном из верхних уровней модели OSI. Шифрованию подлежит только содержательная часть сообщения, которое требуется передать по сети. После зашифрования к ней добавляется служебная информация, необходимая для маршрутизации сообщения, и результат переправляется на более низкие уровни с целью отправки адресату.

Теперь сообщение не требуется постоянно расшифровывать и зашифровывать при прохождении через каждый промежуточный узел сети связи. Сообщение остается зашифрованным на всем пути от отправителя к получателю.

Основная проблема, с которой сталкиваются пользователи сетей, где применяется сквозное шифрование, связана с тем, что служебная информация, используемая для маршрутизации сообщений, передается по сети в незашифрованном виде. Опытный криптоаналитик может извлечь для себя массу полезной информации, зная кто с кем, как долго и в какие часы общается через компьютерную сеть. Для этого ему даже не потребуется быть в курсе предмета общения.

По сравнению с канальным, сквозное шифрование характеризуется более сложной работой с ключами, поскольку каждая пара пользователей компьютерной сети должна быть снабжена одинаковыми ключами, прежде чем они смогут связаться друг с другом. А поскольку криптографический алгоритм реализуется на

верхних уровнях модели OSI, приходится также сталкиваться со многими существенными различиями в коммуникационных протоколах и интерфейсах в зависимости от типов сетей и объединяемых в сеть компьютеров. Все это затрудняет практическое применение сквозного шифрования.

Комбинированное шифрование

Комбинация канального и сквозного шифрования данных в компьютерной сети обходится значительно дороже, чем каждое из них по отдельности. Однако именно такой подход позволяет наилучшим образом защитить данные, передаваемые по сети. Шифрование в каждом канале связи не позволяет противнику анализировать служебную информацию, используемую для маршрутизации. А сквозное шифрование уменьшает вероятность доступа к незашифрованным данным в узлах сети.

При комбинированном шифровании работа с ключами ведется так: сетевые администраторы отвечают за ключи, используемые при канальном шифровании, а о ключах, применяемых при сквозном шифровании, заботятся сами пользователи.

Шифрование файлов

На первый взгляд, шифрование файлов можно полностью уподобить шифрованию сообщений, отправителем и получателем которых является одно и то же лицо, а средой передачи служит одно из компьютерных устройств хранения данных (магнитный или оптический диск, магнитная лента, оперативная память). Однако все не так просто, как кажется на первый взгляд.

Если при передаче по коммуникационным каналам сообщение затеряется по пути от отправителя к получателю, его можно попытаться передать снова. При шифровании данных, предназначенных для хранения в виде компьютерных файлов, дела обстоят иначе. Если вы не в состоянии расшифровать свой файл, вам вряд ли удастся сделать это и со второй, и с третьей, и даже с сотой попытки. Ваши данные будут потеряны раз и навсегда. Это означает, что при шифровании файлов необходимо предусмотреть специальные механизмы предотвращения возникновения ошибок в шифртексте.

Криптография помогает превратить большие секреты в маленькие. Вместо того чтобы безуспешно пытаться запомнить содержимое огромного файла, человеку достаточно его зашифровать и сохранить в памяти использованный для этой цели ключ. Если ключ применяется для шифрования сообщения, то его требуется иметь под рукой лишь до тех пор, пока сообщение не дойдет до своего адресата и не будет им успешно расшифровано. В отличие от сообщений, шифрованные файлы могут храниться годами, и в течение всего этого времени необходимо помнить и держать в секрете соответствующий ключ.

Есть и другие особенности шифрования файлов, о которых необходимо помнить вне зависимости от применяемого криптографического алгоритма:

- нередко после шифрования файла его незашифрованная копия остается на другом магнитном диске, на другом компьютере или в виде распечатки, сделанной на принтере;
- размер блока в блочном алгоритме шифрования может значительно превышать размер отдельной порции данных в структурированном файле, в результате чего зашифрованный файл окажется намного длиннее исходного;

- скорость шифрования файлов при помощи выбранного для этой цели криптографического алгоритма должна соответствовать скоростям, на которых работают устройства ввода/вывода современных компьютеров;
- работа с ключами является довольно непростым делом, поскольку разные пользователи должны иметь доступ не только к различным файлам, но и к отдельным частям одного и того же файла.

Если файл представляет собой единое целое (например, содержит отрезок текста), восстановление этого файла в исходном виде не потребует больших усилий: перед использованием достаточно будет просто расшифровать весь файл. Однако если файл структурирован (например, разделен на записи и поля, как это делается в базах данных), то расшифровывание всего файла цели ком каждый раз, когда необходимо осуществить доступ к отдельной порции данных, сделает работу с таким файлом чрезвычайно неэффективной. Шифрование порций данных в структурированном файле делает его уязвимым по отношению к атаке, при которой злоумышленник отыскивает в этом файле нужную порцию данных и заменяет ее на другую по своему усмотрению.

У пользователя, который хочет зашифровать каждый файл, размещенный на жестком диске компьютера, имеются две возможности. Если он использует ojдин и тот же ключ для шифрования всех файлов, то впоследствии окажется не в состоянии разграничить доступ к ним со стороны других пользователей. Кроме того, в результате у криптоаналитика будет много шифртекста, полученного на одном ключе, что существенно облегчит вскрытие этого ключа.

Лучше шифровать каждый файл на отдельном ключе, а затем зашифровать все ключи при помощи мастер-ключа. Тем самым пользователи будут избавлены от суеты, связанной с организацией надежного хранения множества ключей. Разграничение доступа групп пользователей к различным файлам будет осуществляться путем деления множества всех ключей на подмножества и шифрования этих подмножеств на различных мастер-ключах. Стойкость такой криптосистемы будет значительно выше, чем в случае использования единого ключа для шифрования всех файлов на жестком диске, поскольку ключи, применяемые для шифрования файлов, можно генерировать случайным образом и, следовательно более стойкими против словарной атаки.

Аппаратное и программное шифрование

Аппаратное шифрование

Большинство средств криптографической защиты данных реализовано в виде специализированных физических устройств. Эти устройства встраиваются в линию связи и осуществляют шифрование всей передаваемой по ней информации. Преобладание аппаратного шифрования над программным обусловлено несколькими причинами.

- Более высокая скорость. Криптографические алгоритмы состоят из огромного числа сложных операций, выполняемых над битами открытого текста. Современные универсальные компьютеры плохо приспособлены для эффективного выполнения этих операций. Специализированное оборудование умеет делать их гораздо быстрее.
- Аппаратуру легче физически защитить от проникновения извне. Программа, выполняемая на персональном компьютере, практически беззащитна. Вооружившись отладчиком, злоумышленник может внести в нее скрытые

изменения, чтобы понизить стойкость используемого криптографического алгоритма, и никто ничего не заметит. Что же касается аппаратуры, то она обычно помещается в особые контейнеры, которые делают невозможным изменение схемы ее функционирования. Чип покрывается специальным химическим составом, и в результате любая попытка преодолеть защитный слой этого чипа приводит к самоуничтожению его внутренней логической структуры. И хотя иногда электромагнитное излучение может служить хорошим источником информации о том, что происходит внутри микросхемы, от этого излучения легко избавиться, заэкранировав микросхему. Аналогичным образом можно заэкранировать и компьютер, однако сделать это гораздо сложнее, чем миниатюрную микросхему.

• Аппаратура шифрования более проста в установке. Очень часто шифрование требуется там, где дополнительное компьютерное оборудование является совершенно излишним. Телефоны, факсимильные аппараты и модемы значительно дешевле оборудовать устройствами аппаратной шифрования, чем встраивать в них микрокомпьютеры с соответствующим программным обеспечением.

Даже в компьютерах установка специализированного шифровального оборудования создает меньше проблем, чем модернизация системного программного обеспечения с целью добавления в него функций шифрования данных. В идеале шифрование должно осуществляться незаметно для пользователя. Чтобы добиться этого при помощи программных средств, средства шифрования должны быть упрятаны глубоко в недра операционной системы. С готовой и отлаженной операционной системой проделать это безболезненно не так-то просто. Но даже любой непрофессионал сможет подсоединить шифровальный блок к персональному компьютеру, с одной стороны, к внешнему модему, с другой.

Современный рынок аппаратных средств шифрования информации предлагает потенциальным покупателям 3 разновидности таких средств - самодостаточные шифровальные модули (они самостоятельно выполняют всю работу с ключами), блоки шифрования в каналах связи и шифровальные платы расширения для установки в персональные компьютеры. Большинство устройств первого и второго типов являются узко специализированным и поэтому прежде, чем принимать окончательное решение об их приобретении, необходимо досконально изучить ограничения, которые при установке накладывают эти устройства на соответствующее "железо", операционные системы и прикладное программное обеспечение. А иначе можно выбросит деньги на ветер, ни на йоту не приблизившись к желанной цели. Правда иногда выбор облегчается тем, что некоторые компании торгуют коммуникационным оборудованием, которое уже имеет предустановленную аппаратуру шифрования данных.

Платы расширения для персональных компьютеров являются более универсальным средством аппаратного шифрования и обычно могут быть легко сконфигурированы таким образом, чтобы шифровать всю информацию, которая записывается на жесткий диск компьютера, а также все данные, пересылаемые на дискеты и в последовательные порты. Как правило, защита от электромагнитного излучения в шифровальных платах расширения отсутствует, поскольку нет смысла защищать эти платы, если аналогичные меры не| предпринимаются в отношении всего компьютера.

Программное шифрование

Любой криптографический алгоритм может быть реализован в виде соответствующей программы. Преимущества такой реализации очевидны: программные средства шифрования легко копируются, они просты в использовании, их нетрудно модифицировать в соответствии с конкретными потребностями.

Во всех распространенных операционных системах имеются встроенные средства шифрования файлов. Обычно они предназначены для шифрования отдельных файлов, и работа с ключами целиком возлагается на пользователя. Поэтому применение этих средств требует особого внимания. Во-первых, ни в коем случае нельзя хранить ключи на диске вместе с зашифрованными с их помощью файлами, а во-вторых, незашифрованные копии файлов необходимо удалить сразу после шифрования.

Конечно, злоумышленник может добраться до компьютера и незаметно внести нежелательные изменения в программу шифрования. Однако основная проблема состоит отнюдь не в этом. Если злоумышленник в состоянии проникнуть в помещение, где установлен компьютер, он вряд ли будет возиться с программой, а просто установит скрытую камеру в стене, подслушивающее устройство — в телефон или датчик для ретрансляции электромагнитного излучения — в компьютер. В конце концов, если злоумышленник может беспрепятственно все это сделать, сражение с ним проиграно, даже еще не начавшись.

Сжатие и шифрование

Алгоритмы сжатия данных очень хорошо подходят для совместного использования с криптографическими алгоритмами. Тому есть две причины:

- При вскрытии шифра криптоаналитик, как правило, полагается на избыточность, свойственную любому открытому тексту. Сжатие помогает избавиться от этой избыточности.
- Шифрование данных является весьма трудоемкой операцией. При сжатии уменьшается длина открытого текста, за счет чего сокращается время, которое будет потрачено на его шифрование.

Надо только не забыть сжать файл до того, как он будет зашифрован. После шифрования файла при помощи качественного криптографического алгоритма полученный шифртекст сжать не удастся, поскольку его характеристики будут близки к характеристикам совершенно случайного набора букв. Кстати, сжатие может служить своеобразным тестом для проверки качества криптографического алгоритма: если шифртекст поддается сжатию, значит этот алгоритм лучше заменить на более совершенный.

Модуляция. Амплитудная, частотная и фазовая модуляции. Понятие спектра сигнала. Спектр колебаний модулированного сигнала. Импульсная модуляция.

Способы модуляции.

Модуляция (лат. modulatio – мерность, размерность) – процесс изменения одного или нескольких параметров высокочастотного модулируемого колебания по закону информационного низкочастотного сообщения (сигнала).

В результате спектр управляющего сигнала переносится в область высоких частот, ведь для эффективного вещания в пространство необходимо чтобы все приемопередающие устройства работали на разных частотах и «не мешали» друг

другу. Это процесс «посадки» информационного колебания на априорно известную несущую.

Передаваемая информация заложена в управляющем сигнале. Роль переносчика информации выполняет высокочастотное колебание, называемое несущим. В качестве несущего могут быть использованы колебания различной формы (прямоугольные, треугольные и т.д), однако всего применяются гармонические колебания.

В зависимости от того, какой из параметров несущего колебания изменяется, различают вид модуляции (амплитудная, частотная, фазовая и др.). Модуляция дискретным сигналом называется цифровой модуляцией или манипуляцией.

Амплитудная модуляция

Амплитудная модуляция – вид модуляция, при которой изменяемые параметром несущего сигнала является его амплитуда При амплитудной модуляции на входы модулятора поступают сигнал V и несущая U. Ниже представлена формула амплитудной модуляции.

V=Vm*sin(W*t+j)

Например, если сигнал есть гармоническое колебания с амплитудой Vm, частотой W и фазой j, то на выходе нелинейного элемента в модуляторе будут модулированные колебания.

UAN=Um*(1+msin(W*t+j))*sin(v*t+y)

Где m=Vm/Um – коэффициент модуляции. На выходе модулятора в спектре сигнала присутствуют несущая частота v и две боковые частоты v+W и v-W. Если сигнал занимает некоторую полосу частот, то в спектре модулированного колебания появятся две боковые полосы.

При амплитудной модуляции во избежание искажений, называемых качанием фронта, нужно выполнение условия v>W, где v и W – соответственно несущая и модулирующая частота.

Соблюдение этого условия при стандартной (для среднескоростной аппаратуру передачи данных) несущей частоте 1700Гц не может обеспечить информационные скорости выше 300бит/с.

Поэтому в модемах (рисунок 54) применяют дополнительное преобразование частоты сигнала: производят модуляцию несущей, имеющей повышенную частоты, например Fнд= 10кГц, затем и с помощью фильтра выделяют спектр модулирующие колебания на промежуточную с=частоту, например 1700Гц.

Тогда при боковых полосах до 1400Гц спектр сигнала согласуется с полосой пропускания телефонных линий. Однако достигаемые при этом скорости передачи данных остаются невысокими.

Скорости передачи повышаются с помощью квадратурно-амплитудной или фазовой модуляции за счет того, вместо двоичных модулирующих сигналов используются дискретные сигналы (рисунок 55) с большим числом возможных значений.

Амплитудная манипуляция

Амплитудная манипуляция (АМн; англ. Amplitude shift keying (ASK), а также англ. Continuous wave (CW)) - изменение сигнала, при котором скачкообразно меняется амплитуда несущего колебания. АМн можно

рассматривать как частный случай квадратурной манипуляции (КАМн англ. Quadrature amplitude shift keying(QASK)).

Телеграфные сигналы «азбука Морзе» чаще всего передают при помощи амплитудной манипуляции. В передатчике этот метод реализуется наиболее просто по сравнение с другими видами манипуляции.

Приемник для приема телеграфных сигналов на слух, напротив, несколько усложняется: в нем должен присутствовать сигнал, чтобы на выходе приемника можно было выделить разностную звуковую частоту.

Пригодны приемники прямого преобразования, регенеративные в режиме генерации и супергетеродинные с дополнительным «телеграфным» гетеродином.

Амплитуда высокочастотного сигнала на выходе радиопередатчика принимает только два значения: включения и выключено. Соответственно, включение или выключение («ключевание») выполняется оператором с помощью телеграфного ключе или с помощью автоматического формирователя телеграфных посылок (датчика когда Морзе, компьютера).

Огибающая радиоимпульса (элементарной посылки – точки и тире) на практике, естественно, не прямоугольная (как это показано схематично на рисунке), а имеет плавные передний и задний фронты. В противном случае частотный спектр сигнала может стать недопустимо широким, а при приема сигнала на слух ощущаются неприятные щелчки.

Частотная модуляция

Частотная модуляция (ЧМ) – вид аналоговой модуляции, при котором информационный сигнал управляет частотой несущего колебания. По сравнению с амплитудной модуляцией здесь амплитуда остается постоянной.

Частотная модуляция была предложена американцем Эдвином Армстронгом и запатентована им 26 декабря 1933 года.

Частотная модуляция применяется для высококачественной передачи звукового (низкочастотного) сигнала в радиовещании (в диапазоны УКВ), для звукового сопровождения телевизионных программ, передачи сигналов цветности в телевизионном стандарте SECAM, видеозаписи на магнитную ленту, музыкальных синтезаторах.

Высокое качество кодирования аудиосигнала обусловлено тем, что при ЧМ применяется большая (по сравнению с шириной спектра сигнала АМ) девиация несущего сигнала, а в приемной аппаратуре используют ограничитель амплитуды радиосигнала для ликвидации импульсных помех.

При частотной манипуляции значениям «0» и «1» информационной последовательности соответствуют определенные частоты синусоидального сигнала при неизменной амплитуде. Частотная манипуляция весьма помехоустойчива, поскольку помехи телефонного канала искажают в основном амплитуду, а не частоту сигнала. Однако при частотной манипуляции неэкономно расходуется ресурс полосы частот телефонного канала.

Поэтому этот вид модуляции применяется в низкоскоростных протоколах, позволяющих осуществлять связь по каналам с низким отношением сигнал/шум.

Частотная манипуляция с минимальным сдвигом (англ. Minimal Shift Keying (MSK)) представляет собой способ модуляции, при котором не происходит скачков фазы и изменение частоты происходит в моменты пересечения несущей нулевого уровня.

MSK уникальна потому что значение частот соответствующих логическим «0» и «1» отличаются на величину равную половине скорости передачи данных. Другими словами, индекс модуляции равен 0,5: где , T – длительность бита.

Например, при скорости передачи 1200бит/с MSK-сигнал будет сформирован из колебаний с частотами 1200Гц и 1800Гц соответствующих логическим «0» и «1».

В телеграфировании частотная манипуляция – процесс изменения частоты генератора в соответствии с передающими посылками.

Линейная частотная модуляция

Линейная частотная модуляция (ЛЧМ) сигнала – это вид частотной модуляция, при которой частота несущего сигнала изменяется по линейному закону.

Фазовая модуляция

Фазовая модуляция - один из видов модуляции колебаний, при которой фаза несущего колебания управляется информационным сигналом.

Фазомодулированный сигнал s(t) имеет следующий вид:

$s(t)=g(t)\sin[2\pi fct+\varphi(t)]$

Где g(t) – огибающая сигнала; φ(t) является модулирующим сигналом; fc – частота несущей; t – время.

Фазовая модуляция, не связанная начальной фазой несущего сигнала, называется относительной фазовой модуляцией (ОФМ).

В случае, когда информационный сигнал является дискретным, то говорят о фазовой манипуляции.

Хотя, строго говоря, в реальных изделиях манипуляции не бывает, так как для сокращения занимаемой полосы частот манипуляция производится не прямоугольным импульсом, а колоколообразным (приподнятым косинусом и др.).

Несмотря на это, при модуляции дискретным сигналом говорят только о манипуляции.

Фазовая модуляция (PSK – Phase Shift Keying) двумя уровнями сигнала (1 и 0) осуществляется переключением между двумя несущими, сдвинутыми на полпериода друг относительно друга (рисунок 65). Другой вариант PSK- изменение фазы на p/2 в каждом такте при передаче нуля и на ¾*p, если передается единица.

Фазовая манипуляция (ФМн, англ. Phase-shift keying (PSK)) – один из них видов фазовой модуляции, при которой фаза несущего колебания меняется скачкообразно.

Модулирующий сигнал, несущая и фазоманипулированный сигнал системы спутниковой навигации NAVSTAR GPS.

Двоичная фазовая манипуляция (ФМн, англ. BPSK – binary phase-shift keying) – самая простая форма фазовой манипуляции. Работа схемы двоичной ФМн заключается в смещении фазы несущего колебания на одно из двух значений, нуль или π(180o).

При квадратурной фазовой манипуляции (англ. QPSK – Quardrature Phase Sfift Keying) используется созвездие из четырех точек, размещенных на равных расстояниях по окружности.

Используя 4 фазы, в QPSK на символ приходится два бита, как показано на рисунке. Анализ показывает, что скорость может быть увеличена в два раза

относительно BPSK при той же полосе сигнала, либо оставить скорость прежней, но уменьшить полосу вдвое.

Существуют фазовые манипуляции более высоких уровней – восьмеричная ФМн и др.

Однополосная модуляция SSB

Однополосная модуляция SSB (Амплитудная модуляция с одной боковой полосой) (ОМ, англ. Single-sideband modulion, SSB) – разновидность амплитудной модуляции (АМ), широко применяемая в аппаратуре каналообразования для эффективного использования спектра канала и мощности передающей радиоаппаратуры.

Для передачи несущей частоты обычного радиосигнала с АМ используется часть мощности передающей аппаратуре (около 50%), поэтому отсутствие в ОМ – сигнале сигнала несущей частоты, а также одной из боковых полос дает возможность использовать всю мощность передающей аппаратуры для передачи только полезного сигнала .

Недостатком метода являются высокие требования к частотной точности приемной аппаратуры. Модуляция SSB ввиду своей эффективности, широко используется в профессиональной и любительской радиосвязи на коротких и ультракоротких волнах.

Квадратурно - амплитудная модуляция

Квадратурно - амплитудная модуляция (КАМ) (QAM – Quadrature Amplitude Modulation, ее также называют квадратурно-импульсной) – разновидность амплитудной модуляции сигнала, которая представляет собой сумму двух несущих колебаний одной частоты, но сдвинутых по фазе относительно друг друга на 90 градусов, каждая из которых модулирована по амплитуде своим модулирующим сигналом.

Модуляция основана на передаче одним элементом модулированного сигнала n бит информации, где n=4…8(т.е. используются 16…256 дискретных значений амплитуды). Однако для надежного различения этих значений амплитуды требуется малый уровень помех (отношение сигнал/помеха не менее 12 дБ при n=4).

При меньших отношениях сигнал/помеха лучше применять фазовую модуляцию с четырьмя или восемью дискретными значениями фазы для представления соответственно 2 или 3 бит информации.

Тогда при скорости модуляции в 1200 бод (т.е. 1200 элементов аналогового сигнала в секунду, где элемент – часть сигнала между возможными сменами фаз) и четырехфазной модуляции скорость передачи данных равно 2400 бит/с.

Используются также скорости передачи 4800 бит/с (при скорости 1600 бод и восьмифазной модуляции), 9600 бит/с и более при комбинации фазовой и амплитудной модуляций.

Кодово-импульсная модуляция

Кодово-импульсная модуляция (КИМ или PCM – Pulse Code Modula-tion) используются для передачи аналоговых сигналов по цифровым каналам связи.

Этот вид модуляции сводится к измерению амплитуды аналогового сигнала в моменты времени, отстоящие друг от друга на dt, и кодированию этих амплитуд цифровым кодом.

Величина dt определяется по теореме Котельникова: для неискаженной передачи нужно иметь не менее двух отсчетов на период колебаний, соответствующий высшей составляющей в частотном спектре сигнала.

В цифровых каналах ISDN (Integrated Services Digital Network) за основу принята передача голоса с частотным диапазоном до 4кГц, а кодирование производится восемью (или семью) битами. Отсюда получаем, что частота отсчетов (передачи байтов) равно 8кГц, т.е. биты передаются с частотой 64кГц (или 56кГц, при семибитовой кодировке).

При преобразовании амплитуды А аналогового сигнала в цифровой код К желательно учитывать нелинейность амплитудных характеристик приборов и иметь зависимость К от А монотонно убывающей с ростом амплитуды.

Разновидностями КИМ являются дельта-модуляция (ДМ), дифференциальная ДМ (ДДМ) и адаптивная ДМ (АДДМ). В них передаются разности амплитуд А1 и А2 соседних отсчетов.

При этом ДМ А1 – амплитуда на входе модулятора, и А2 – амплитуда отсчета, которая соответствует переданному сигналу, в предыдущем временном такте.

Для представления разности используется всего 1 бит (т.е. передается знак разности), поэтому нужна достаточно высокая частота отсчетов, чтобы не было «запаздывания» изменений передаваемого сигнала по сравнению с реальными изменениями.

ДДМ отличается от ДМ тем, что знак разности А1- А2 передается только в момент пересечения величиной А1 одного из уровней квантования. В АДДМ шаги отсчетов адаптируются к динамики изменения величины сигнала.

Характеристика модуляции

В сравнительно простых модемах применяют частотную модуляцию (FSK – Frequency Shift Keying) со скоростями передачи до 1200бит/с. Так, если необходима дуплексная связь по двухпроводной линии, то возможно представление 1 и 0 в вызывном модеме частотами 980 и 1180 Гц соответственно, а в ответном модеме – 1650 и 1850 Гц. При этом скорость передачи составляет 300 бод.

Обычно для передачи сигнала об ошибке от приемника к передатчику нужен канал обратной связи. При этом требования к скорости передачи данных по обратному каналу могут быть невысокими.

Тогда в полосе частот телефонного канала образуют обратный канал с ЧМ, по которому со скоростью 75 бит/с передают 1 частотой 390 Гц и 0 частотой 450 Гц.

По характеристикам фазовая модуляция близка к частотной модуляции. В случае синусоидального модулирующего (информационного) сигнала, результаты частотной и фазовой модуляции совпадают.

Тема 5.3. ПЕРЕДАЧА ИНФОРМАЦИИ

Передача информации и коммуникационные технологии. Коммуникационные сети. Принципы построения цифровых каналов. Информационная модель канала связи.

Пропускная способность дискретного канала без шума. Основная теорема Шеннона для дискретного канала без шума. Эффективное кодирование. Коды Шеннона-Фано и Хаффмена. Современные методы сжатия данных. Пропускная способность дискретного канала с шумом. Основная теорема Шеннона для дискретного канала с шумом. Пропускная способность непрерывного канала с шумом.

Методы повышения помехоустойчивости передачи данных. Помехи. Модели ошибок в реальных каналах. Основные методы повышения помехоустойчивости передачи данных. Методы оптимального приема сигналов. Бинарное обнаружение. Критерии оптимальности бинарного обнаружения. Структура оптимального приемника.

Помехоустойчивое кодирование. Принципы помехоустойчивого кодирования. Принципы построения корректирующих кодов. Понятие группы и поля. Групповые коды. Коды Хэмминга. Порождающая и проверочная матрицы групповых кодов. Циклические коды. Принципы повышения помехоустойчивости передачи данных, основанные на использования обратной связи между выходом и входом канала.

Способы контроля правильности передачи информации

Основные сведения о кодировании

При передаче информации по некачественным и/или разделяемым каналам связи возможны ошибки, то есть искажения передаваемой информации (рисунок 1). Эти ошибки необходимо выявлять и исправлять.

Рисунок 1 – Помехи при передаче информации

Управление правильностью (помехозащищенностью) передачи информации (рисунок 2) выполняется с помощью помехоустойчивого кодирования.

Кодирование это представление сообщения последовательностью элементарных символов. Различают коды, обнаруживающие ошибки, и корректирующие коды, которые дополнительно к обнаружению еще и исправляют ошибки.

Рисунок 2 – Схема передачи информации

Помехозащищенность достигается с помощью введения избыточности. Устранение ошибок с помощью корректирующих кодов (такое управление называют Forward Error Control) реализуют в симплексных каналах связи.

В дуплексных каналах достаточно применения кодов, обнаруживающих ошибки (Feedback or Backward Error Control), так как сигнализация об ошибке вызывает повторную передачу от источника. Это основные методы, используемые в информационных сетях.

Простейшими способами обнаружения ошибок являются контрольное суммирование, проверка на нечетность. Однако они недостаточно надежны, особенно при появлении пачек ошибок.

Поэтому в качестве надежных обнаруживающих кодов применяют циклические коды. Примером корректирующего кода является код Хемминга. Пример изображен на рисунке 3.

Рисунок 3 – Пример кода Хемминга

Рассмотрим кодирование дискретных сообщений. Символы в сообщениях могут относиться к алфавиту, включающему n букв (буква – символ сообщения).

Однако число элементов кода k существенно ограничено сверху энергетическими соображениями, т.е. часто $n > k$.

Так, если отношение сигнал/помеха для надежного различения уровня сигнала должно быть не менее q,то наименьшая амплитуда для представления одного из k символов должна быть q*g, где g – амплитуда помехи, а наибольшая амплитуда соответственно q*g*k.

Мощность передатчика пропорциональна квадрату амплитуды сигнала (тока или напряжения), т.е. должна превышать величину, пропорциональную (q*g*k).

В связи с этим распространено двоичное кодирование (рисунок 113) с k =2. При двоичном кодировании сообщений с n типам букв, каждая из n ,букв кодируется определенной комбинацией 1 и 0 (например, код ASCII). Контроль правильности передачи информации представлен на рисунке 4.

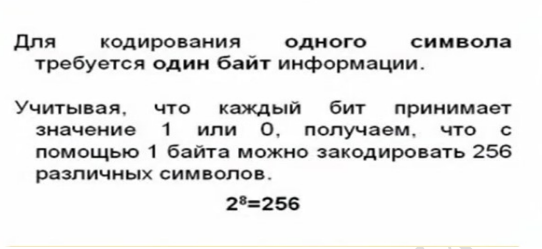


Рисунок 3 – Двоичное кодирование

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00	00 □	01 □	02 □	03 □	04 □	05 □	06 □	07 □	08 □	09 □	0A □	0B □	0C □	0D □	0E □	0F □
10	10 □	11 □	12 □	13 □	14 □	15 □	16 □	17 □	18 □	19 □	1A □	1B □	1C □	1D □	1E □	1F □
20	20	21 !	22 “	23 #	24 $	25 %	26 &	27 (	28)	29 '	2A *	2B +	2C ,	2D -	2E .	2F /
30	30 0	31 1	32 2	33 3	34 4	35 5	36 6	37 7	38 8	39 9	3A :	3B ;	3C <	3D =	3E >	3F ?
40	40 @	41 A	42 B	43 C	44 D	45 E	46 F	47 G	48 H	49 I	4A J	4B K	4C L	4D M	4E N	4F O
50	50 P	51 Q	52 R	53 S	54 T	55 U	56 V	57 W	58 X	59 Y	5A Z	5B [	5C \	5D]	5E ^	5F _
60	60 `	61 a	62 b	63 c	64 d	65 e	66 f	67 g	68 h	69 i	6A j	6B k	6C l	6D m	6E o	6F p
70	70 p	71 q	72 r	73 s	74 t	75 u	76 v	77 w	78 x	79 y	7A z	7B {	7C |	7D }	7E ~	7F □

Рисунок 4 – Контроль правильности передачи информации

Кодирование аналоговых сообщений после их предварительной дискретизации должно выполняться в соответствии с теоремой Котельникова: если в спектре (рисунок 5) функции f(t) нет частот выше F,то это функция может быть полностью восстановлена по совокупности своих значений, определенных в моменты времени t, отстоящие друг от друга на величину 1/(2*Fв).

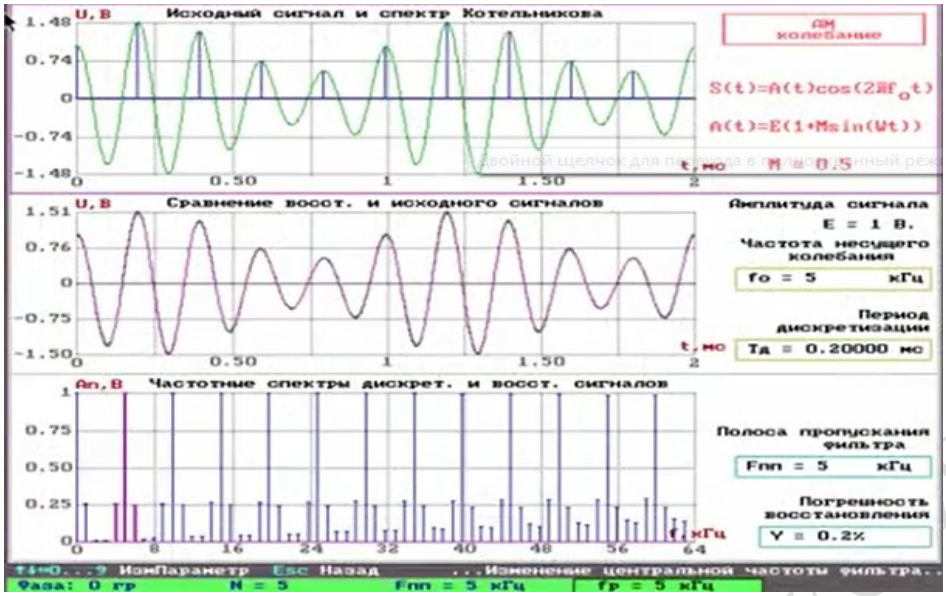


Рисунок 5 – Спектр Котельникова

Для передачи аналогового сигнала производится его дискретизация с частотой отсчетов 2*Fв и выполняется кодово–импульсная модуляция последовательности отсчетов.

Количество информации в сообщении (элементе сообщения) определяется по формуле:

$$I = -\log_2 P$$

Где P – вероятность появления сообщения (элемента сообщения). Из этой формулы следует, что единица измерения количества информации есть количество информации, содержащееся в одном бите двоичного кода при условии равной вероятности появления в нем 1 и 0.

В то же время один разряд десятичного кода содержит $I=-\log_2 P=3{,}32$ единиц информации (при том же условии равно вероятности появления десятичных символов. Т.е. при P=0,1).

Энтропия (рисунок 6) источника информации с независимыми и равновероятными сообщениями есть среднее арифметическое количеств информации сообщений:

$$H = -\sum P_k \log_2 P_k, \text{ где } k=1\ldots N$$

Где Pk – вероятность появления сообщения. Другими словами, энтропия есть мера неопределенности ожидаемой информации.

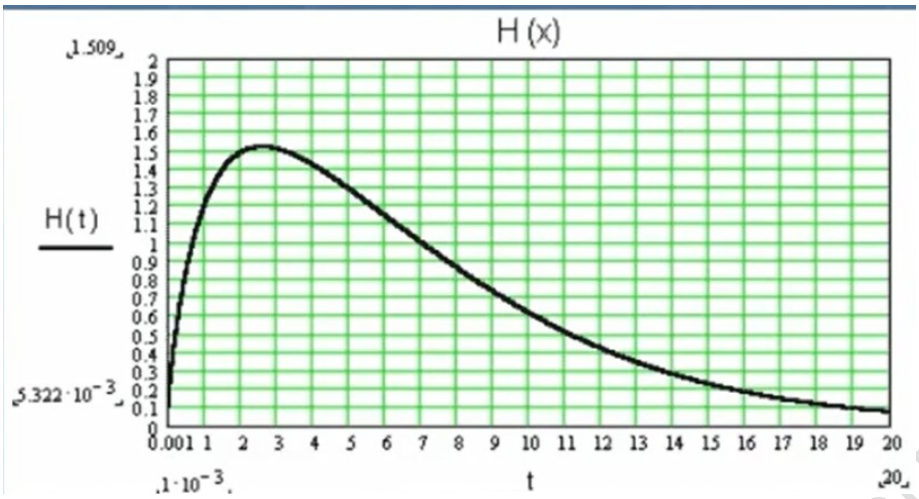


Рисунок 6 – Энтропия

Например, пусть имеем два источника информации (рисунок 7), один передает двоичный код с равновероятным появлением в нем 1 и 0, другой имеет вероятность 1, равную 2, и вероятность 0, равную 1–2. Очевидно, что неопределенность получения в очередном такте символа 1 или 0 от первого источника выше, чем от второго. Это подтверждается количественно оценкой энтропии: у первого источника H=1, у второго приблизительно H=–2*log2, т.е. значительно меньше.

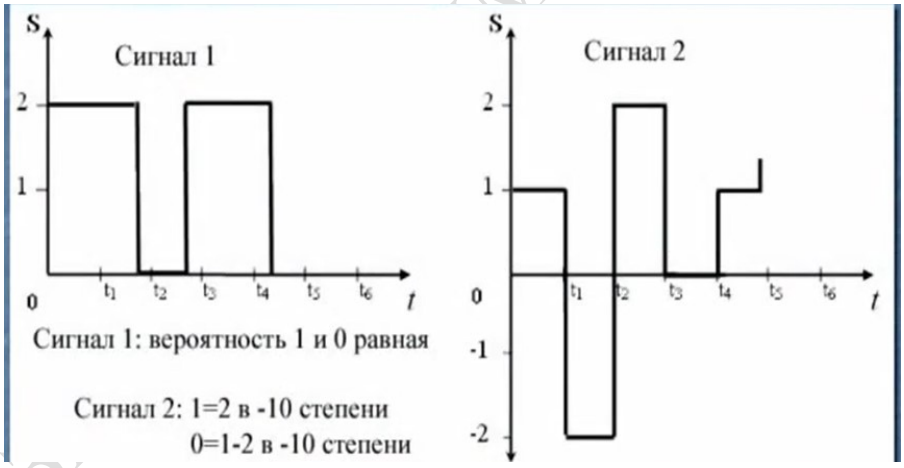
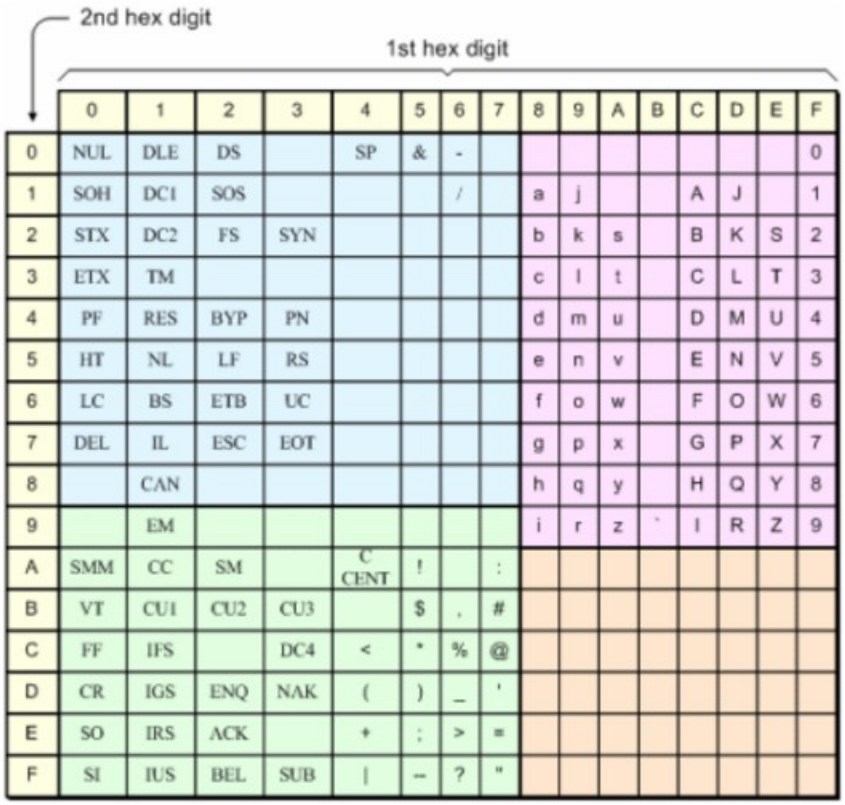


Рисунок 7 – Два источника информации

Основные используемые коды

Широко используемые двоичные коды:

EBCDIC (Extended Binary Coded Decimal Interchange Code) – символы кодируются восемью битами; популярен благодаря его использованию в IBM(рисунок 8);



2nd hex digit \ 1st hex digit	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	DS		SP	&	-									0
1	SOH	DC1	SOS				/		a	j			A	J		1
2	STX	DC2	FS	SYN					b	k	s		B	K	S	2
3	ETX	TM							c	l	t		C	L	T	3
4	PF	RES	BYP	PN					d	m	u		D	M	U	4
5	HT	NL	LF	RS					e	n	v		E	N	V	5
6	LC	BS	ETB	UC					f	o	w		F	O	W	6
7	DEL	IL	ESC	EOT					g	p	x		G	P	X	7
8		CAN							h	q	y		H	Q	Y	8
9		EM							i	r	z	`	I	R	Z	9
A	SMM	CC	SM		C CENT	!		:								
B	VT	CU1	CU2	CU3		$	,	#								
C	FF	IFS		DC4	<	*	%	@								
D	CR	IGS	ENQ	NAK	(	)	_	'								
E	SO	IRS	ACK		+	;	>	=								
F	SI	IUS	BEL	SUB	|	¬	?	"								

Рисунок 8 – Таблица EBCDIC

ASCII (American Standards Committee for Information Interchange) – семибитовый двоичный код (рисунок 9).

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

Рисунок 9 – Таблица ASCII

Оба этих кода включают битовые комбинации для печатаемых символов и некоторых распространенных командных слов типа NUL,CR,ACK,NAK и др.

Для кодировки русского текста нужно вводить дополнительные битовые комбинации. Семибитовая кодировка здесь уже недостаточна.

В восьмибитовой кодировке нужно под русские символы отводить двоичные комбинации, не занятые в общепринятом коде, чтобы сохранять неизменной кодировку латинских букв и других символов.

Так возникли кодировка КОИ8 (рисунок 10), затем при появлении персональных ЭВМ – альтернативная кодировка и при переходе к Windows – кодировка 1251. Множество используемых кодировок существенно усложняет проблему согласования почтовых программ в глобальных сетях.

	.0	.1	.2	.3	.4	.5
C.	ю 44E	а 430	б 431	ц 446	д 434	е 435
D.	п 43F	я 44F	р 440	с 441	т 442	у 443
E.	Ю 42E	А 410	Б 411	Ц 426	Д 414	Е 415
F.	П 41F	Я 42F	Р 420	С 421	Т 422	У 423

	.6	.7	.8	.9	.A	.B
C.	ф 444	г 433	х 445	и 438	й 439	к 43A
D.	ж 436	в 432	ь 44C	ы 448	з 437	ш 448
E.	Ф 424	Г 413	Х 425	И 418	Й 419	К 41A
F.	Ж 416	В 412	Ь 42C	Ы 428	З 417	Ш 428

	.C	.D	.E	.F
C.	л 43B	м 43C	н 43D	о 43E
D.	э 44D	щ 449	ч 447	ъ 44A
E.	Л 41B	М 41C	Н 41D	О 41E
F.	Э 42D	Щ 429	Ч 427	Ъ 42A

Рисунок 10 – Таблица КОИ8–R

Асинхронное и синхронное кодирование

Для правильного распознавания позиций символов в передаваемом сообщении получатель должен знать границы передаваемых элементов сообщения. Для этого необходима синхронизация передатчика и приемника.

Использование специального дополнительного провода для сигналов синхронизации (в этом случае имеем битовую синхронизацию) слишком дорого, поэтому используют другие способы синхронизации.

В асинхронном режиме (рисунок 11) применяют коды. В которых явно выделены границы каждого символа (байта) специальными стартовым и стоповым символами.

Подобные побайтно выделенные коды называют байт – ориентированными, а способ передачи – байтовой синхронизацией. Однако это увеличивает число битов, не относящихся собственно к сообщению.

В синхронном режиме (рисунок 11) синхронизм поддерживается во время передачи всего информационного блока без обрамления каждого байта. Такие коды называют бит – ориентированными.

Для входа в синхронизм нужно обозначать границы лишь всего передаваемого блока информации с помощью специальных начальной и

конечной комбинаций байтов (обычно это двухбайтовые комбинации). В этом случае синхронизация называется блочной (фреймовой).

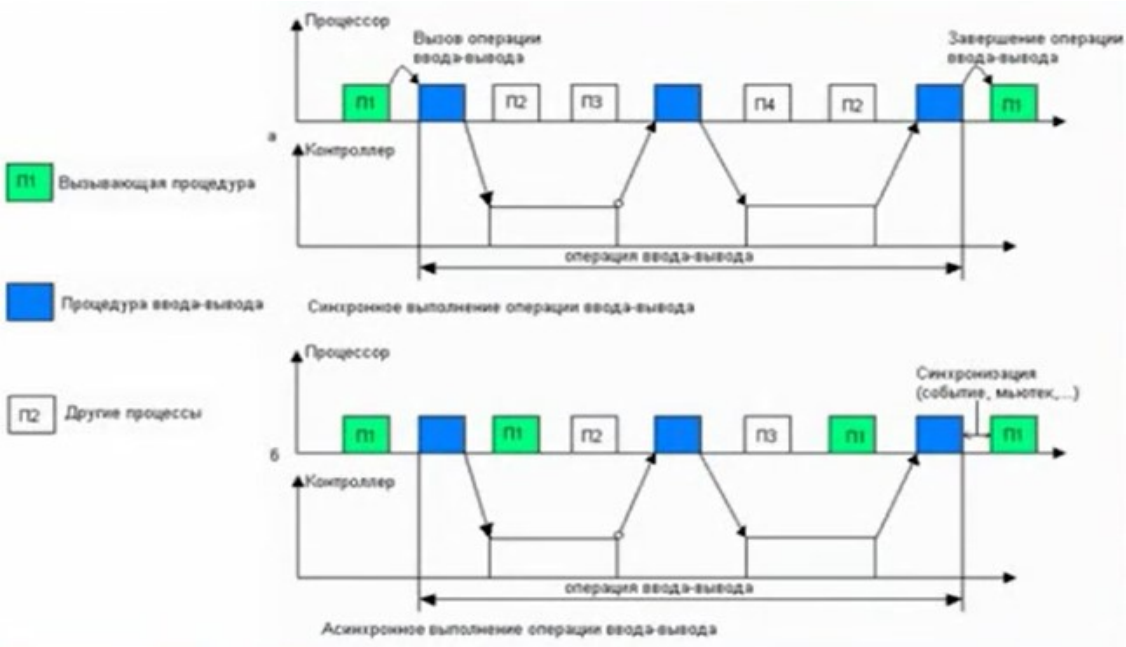


Рисунок 11 – Асинхронное и синхронное кодирование

Для обрамления текстового блока (текст состоит только из печатаемых символов) нужно использовать символы, отличающиеся от печатаемых.

Для обрамления двоичных блоков применяют специальный символ (обозначим его DLE), который благодаря стаффингу становится уникальным. Уникальность заключается в этом, что если DLE встречается внутри блока, то сразу вслед за ним вставляется еще один DLE.

Приемник будет игнорировать каждый второй идущий подряд символ DLE. Если же DLE встречается без добавления, то это граница блока.

Манчестерское кодирование

Передаваемые данные представляются электрическими сигналами. Возможны коды RZ (Return–to–zero), использующие двуполярные сигналы для изображения 1 и 0, и коды NRZ (non–return–to–zero) – коды без возвращения к нулю.

Для кодирования информации наибольшее распространение получили самосинхронизирующиеся коды, так как при этом отпадает необходимость иметь дополнительную линую для передачи синхросигналов между узлами сети.

В ЛВС чаще других применяют манчестерский код, один из разновидностей которого приведена на рисунке 122. Самосинхронизация обеспечивается благодаря формированию синхроимпульсов из перепадов, имеющихся в каждом такте манчестерского кода.

Представления на рисунке 122 разновидность манчестерского кода используется при байт– ориентированном кодировании, при котором каждый байт, состоящий из 1 и 0, обрамляется символами j и k.

В этом случае станция, получившая полномочия, начинает передавать сери сигналов jkjkjk… для того, чтобы станция – получатель могла войти в синхронизм с передающей станцией.

После нескольких jk начинают передаваться байты самого сообщения. Различение четырех возможных значений сигнала выполняется в соответствии с правилами кодирования, представленными в нижней части рисунка 13.

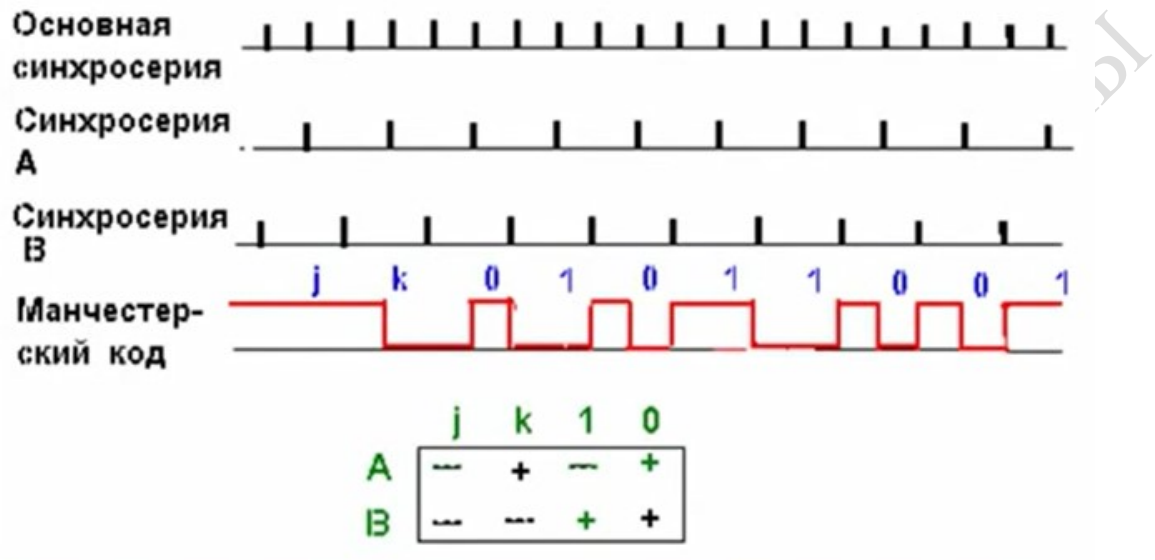


Рисунок 13 – Манчестерское кодирование

Код Хемминга

В коде Хемминга (рисунок 14) вводится понятие кодового расстояния d (расстояния между двумя кодами), равного числу разрядов с неодинаковыми значениями. Возможности исправления ошибок связаны с минимальными кодовым расстоянием dmin.

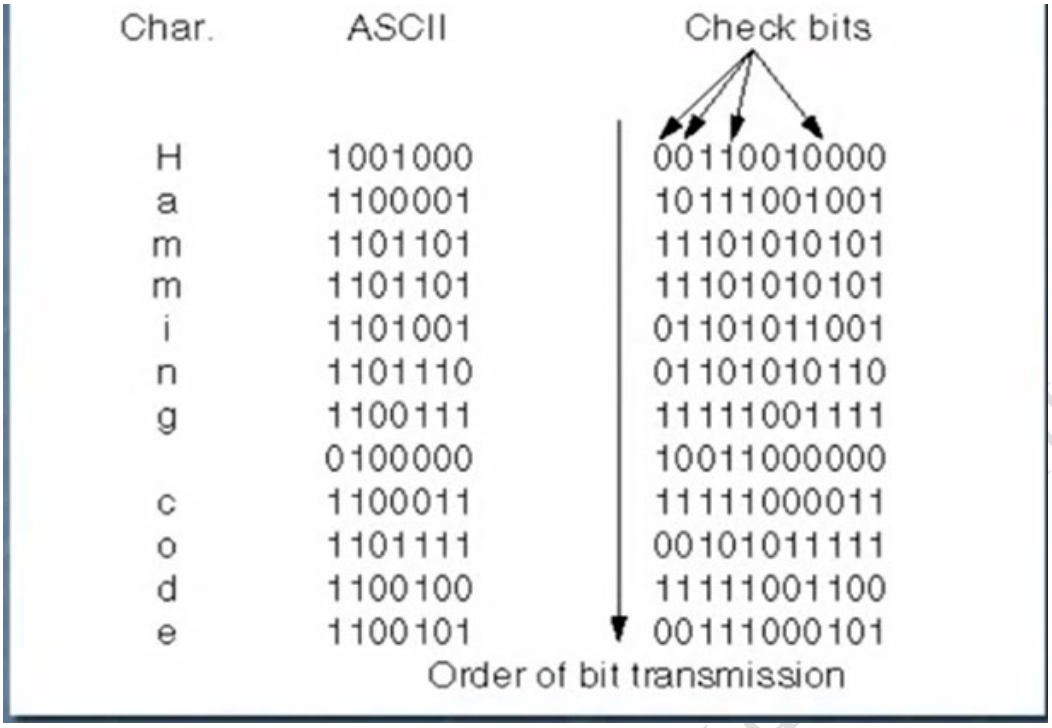


Рисунок 14 – Код Хемминга

Теперь вместе с основным кодом будет передан и дополнительный. На приемном конце вновь рассчитывается дополнительный код и сравнивается с переданным.

Фиксируется код сравнения (поразрядная операция отрицания равнозначности). И если он отличен от нуля, от его значение есть номер ошибочно принятого разряда основного кода.

Так, если принят код 100010, то рассчитанный в приемнике дополнительный код равен инверсии от 010 # 110 = 100, т.е. 011, что означает ошибку в 3–м разряде.

Циклические коды

К числу эффективных кодов, обнаруживающих одиночные, кратные ошибки и пачки ошибок, относятся циклические коды (CRC – Cyclic Redundance Code). Они высоконадежны и могут применяться при блочной синхронизации, при которой выделение, например, бита нечетности было бы затруднительно.

Один из вариантов циклического кодирования заключается в умножении исходного кода на образующий полином g(x), а декодирование – в делении на g(x). Если остаток от деления не равен нулю, то произошла ошибка. Сигнал об ошибке поступает на передатчик, что вызывает повторную передачу.

Образующий полином есть двоичное представление одного из простых множителей, на которые раскладывается число Xn–!, где Xn обозначает единицу в n-м разряде, n равно числу разрядов кодовой группы. Так, если n = 10 и X = 2, то Xn –1 = 1023 = 11*93, и если g(X) = 11 или в двоичном коде

1011, то примеры циклических кодов Aig(X) чисел Ai, в кодовой группе при этом образующем полиноме можно видеть на рисунке 15.

Число	Циклический код	Число	Циклический код
0	0000000000	13	0010001111
1	0000001011	14	0010011010
2	0000010110	15	0010100101
3	0000100001	16	0011000110
5	0000110111	18	0011000110
6	0001000010	19	0011010001

Рисунок 15 – Примеры циклических кодов

Основной вариант циклического кода, широко применяемый на практике, отличается от предыдущего тем, что операция деления на образующий полином заменяется следующим алгоритмом:

1) к исходному кодируемому числу А справа приписывается К нулей, где К – число битов в образующем полиноме, уменьшенное на единицу;
2) над полученным числом А*2К выполняется операция О, отличающаяся от деления тем, что на каждом шаге операции вместо вычитания выполняется поразрядная операция "исключающее ИЛИ";
3) полученный остаток В и есть CRC - избыточный К-разрядный код, который заменяет в закодированном числе С приписанные справа К нулей, т.е. :

$$C = A*2K + B.$$

На приемном конце, над кодом С выполняется операция О. Если остаток не равен нулю, то при передаче произошла ошибка и нужна повторная передача кода А.

Пример. Пусть А = 1001 1101, образующий полином 11001.

Так как К = 4, то А*2К = 100111010000. Выполнение операции О расчета циклического кода изображено на рисунке 16.

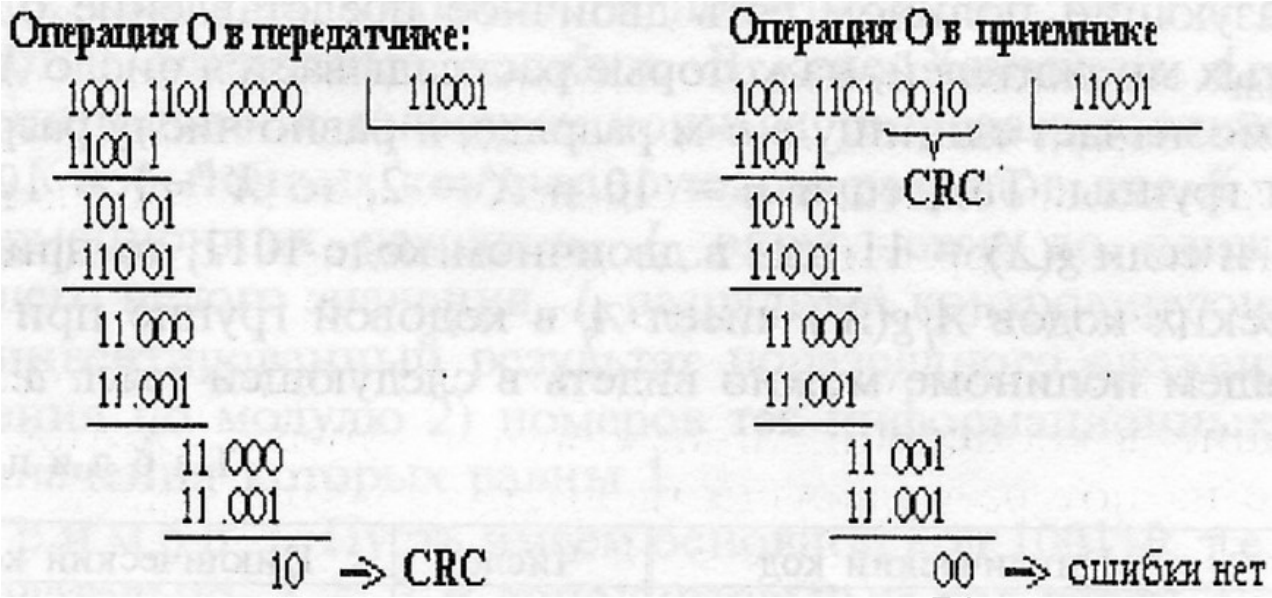


Рисунок 16 – Расчет циклического кода

Положительными свойствами циклических кодов являются малая вероятность необнаружения ошибки и сравнительно небольшое число избыточных разрядов.

Общепринятое обозначение образующих полиномов дает следующий пример: $g(X) = X^{16} + X^{12} + X^5 + 1$, что эквивалентно коду 1 0001 0000 0010 0001. Этот полином используется в протоколе V.42 для кодирования кодовых групп в 240 разрядов с двумя избыточными байтами. В этом протоколе возможен и образующий полином для четырех избыточных байтов:

$$g(X) = X^{32} + X^{26} + X^{23} + X^{22} + X^{16} + X^{12} + X^{11} + X^{10} + +X^8+X^7+X^5+X^4+X^2+1.$$

Алгоритмы сжатия данных

Общие сведения о сжатии

Сжатие данных (графических изображений, видеоизображений и звука) - процедура их перекодирования, производимая с целю уменьшения их объема. Применяется для более рационального использования устройств хранения и передачи данных. Алгоритмы сжатия основаны на устранении избыточности информации, содержащейся в исходных данных (рисунок 17).

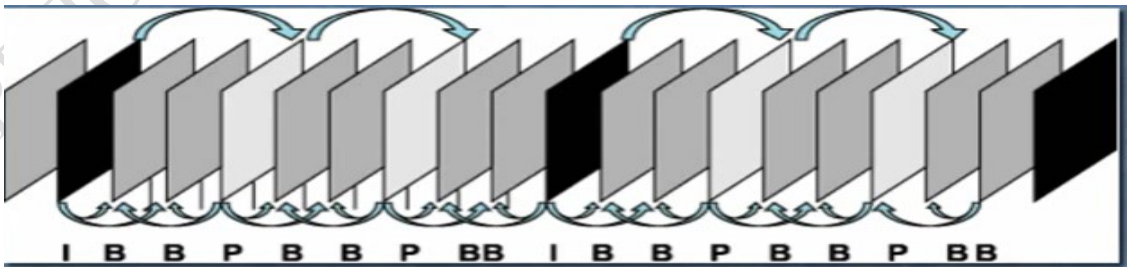


Рисунок 17 – Работа алгоритма сжатия

Коэффициент избыточности сообщения А определяется по формуле: r= (Imax-I)/Imax,где I-количества информации в сообщении А возможное количество информации в сообщении той же длины, что и А

Пример избыточности дают сообщения на естественных языках так, у русского языка r находится в переделах 0,3…0.5

Подобная избыточность обычно устраняется заменой повторяющейся последовательности более коротким значение (кодом).

Другой вид избыточности связан с тем, что некоторые значения в сжимаемых данных встречаются чаще других, при зетом возможно заменять часто встречающиеся данные более короткими кодами, а редкие - более длинами (вероятностное сжатие).

Сжатие данных, не обладающих свойством избыточности (например, случайный сигнал или шум), невозможно без патеры. Также, обычно невозможно сжатие зашифрованной информации.

Наличные сообщениях избыточности позволяет ставать вопрос о сжатии данных, т.е. о передаче того же количества информации с помощью последовательностей символов меньшей длины.

Для этого используются специальные алгоритмы сжатия (рисунок 18), уменьшающие избыточность.

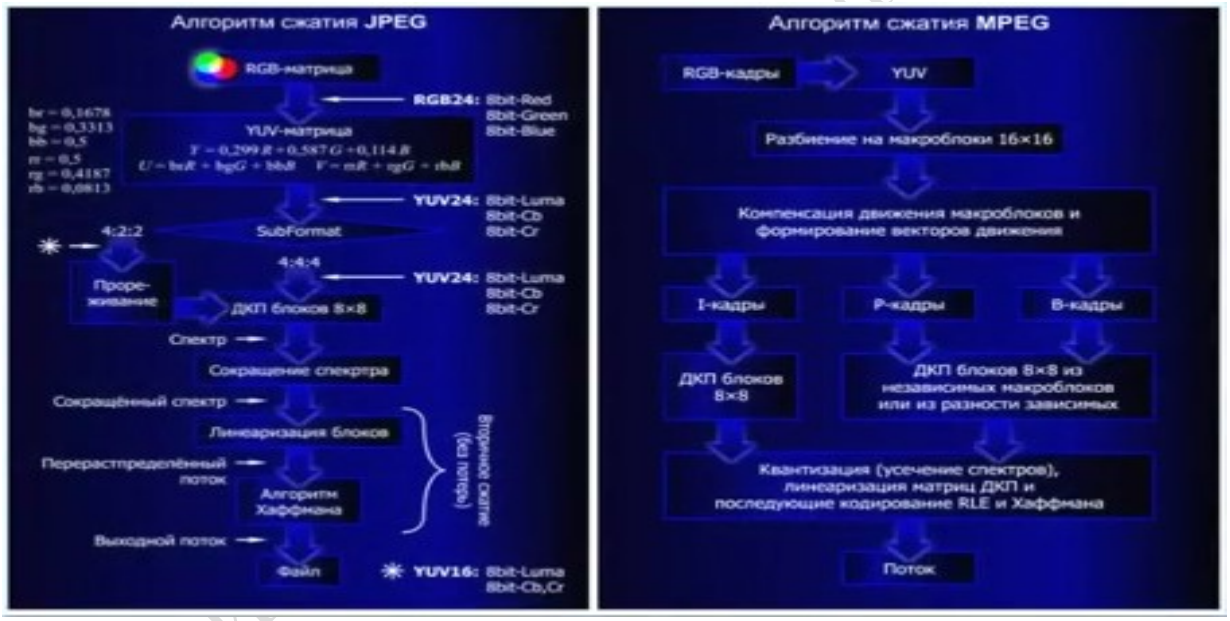


Рисунок 18 – Различные алгоритмы сжатия

Эффект сжатия оценивают коэффициентом сжатия К=n/q где n-число минимально необходимых символов для передачи сообщения (практически это число символов на выходе эталонного алгоритма сжатия);q-число символов в сообщении, сжатом данным алгоритмом. Так, при двоичном кодировании н равно энтропии источника информации.

Наряду с методами сжатия, не уменьшающими количества информации в сообщении, применяются методы сжатия, основанные на потере малосущественной информации.

Алгоритмы сжатия делятся на сжатие без потерь, и сжатие с потерями.

Сжатие без потерь: возможно восстановление исходных данных без искажений, используется при обработке компьютерных программы и данных, реже – для сокращения объема звуковой, фото и видеоинформации.

Наиболее известные алгоритмы сжатия без потерь:

Преобразование Берроуза – Уиллера; преобразование Киндера; алгоритм DEFLATE; дельта-кодирование; энтропионное кодирование; инкрементное кодирование; Алгоритмы Лемпеля-Зива и другие.

Сжатие с потерями: восстановление возможно с искажениями несущественными с точки зрения дальнейшего использование восстановленных данных. Применяться для сокращения объема звуковой, фото видеоинформации, оно значительно эффективнее сжатия без потерь.

Наиболее известные алгоритмы сжатия с потерями: JPEG; линейное предсказывающее кодирование; А - закон; Мою - закон; фрактальное сжатие; трансформирующее кодирование; векторная квантизация; ветвленное сжатие.

Сжатие данных осуществляется либо на прикладном уровне с помощью программ сжатия, таких как ARJ, ZIP, RAR (рисунок 19), либо с помощью устройств защищают ошибок (УЗО) непосредственно в составе модемов, по протоколам типа V.42bis.

Рисунок 19 – Программы сжатия данных

Очевидный способ сжатия числовой информации, представленной в коде ASCII, заключается в использовании сокращенного кода с четырьмя битами на символ вместо восьми так как передается набор включающий только 10 цифр, символы точка, запятая и пробел.

Среди простых алгоритмов сжатия наиболее известны алгоритмы RLE (Run Length Encoding). В них вместо передачи цепочки из одинаковых символов передаются символ и значение длины цепочки. Метод эффективен при передаче растровых изображений, но малополезен при передаче текста.

К методам сжатия относят также методы разностного кодирования, поскольку разности амплитуд отчества представляются меньшим числом разрядов, чем сами амплитуды. Разностное кодирование реализовано в методах дельта модуляции и ее разновидностях.

К методам сжатия относят также методы разностного кодирования поскольку разности амплитуд отсчетов представляются меньшим числом разрядов чем сами амплитуды.

Разностное кодирование реализовано в методах дельта – модуляции и ее разновидностях.

Предсказывающие (предикативные) методы основаны на экстраполяции значений амплитуд отсчетов и если выполнено условие Ar-Ap>d то отсчет должен быть передан иначе он является избыточным, здесь Ar и Ap-амплитуды реального и предсказанного отчетов. d-допуск (допустимая погрешность представления амплитуд). Иллюстрация предсказывающего метода с линейной экстраполяций представлена рисунке 129. Здесь точками показаны предсказываемые значения сигнала. Если точка выходит за пределы передачи отсчета нет, то на приемном конце принимается экстраполированное.коридора (допуска д), показаного пунктирными линиями то пройсходит передача отсчета.

На рисунке 20 передаваемые отсчеты отмечены темными кружками в моменты времени t1, t2, t4, t7.

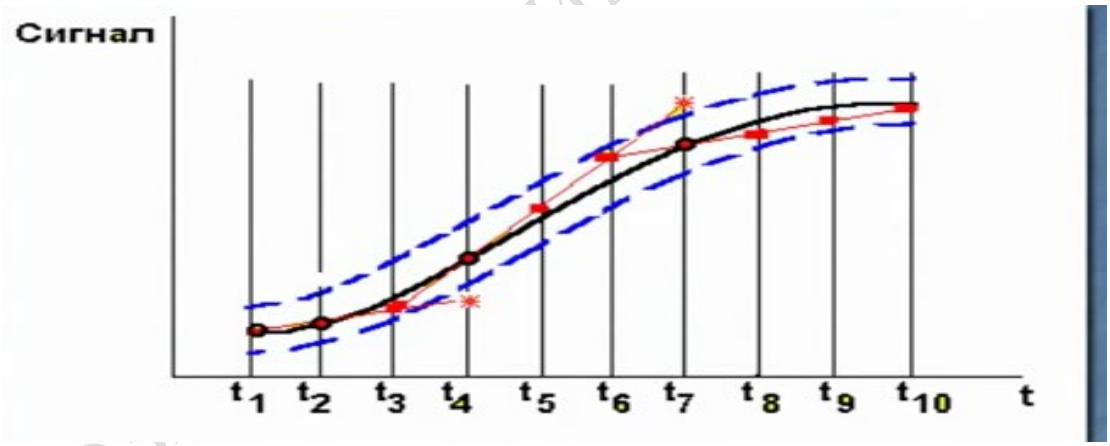


Рисунок 20 – Разность амплитуд

Алгоритмы MPEG (рисунок 21) становится мировыми стандартами для цифрового телевидения. Для сжатия данных об изображениях можно использовать также методы типа (Joint photographic Expert), основанные на потере малосущественной информации (не различимые для глаза оценки кодируются одинаково коды могут стать короче).

В этих методах передаваемая последовательность пикселей делится на блоки в каждом блоке производиться преобразование Фуре устраняются высокие частоты передаются коэффициенты разложения для оставшихся частот, по ним в приемнике изображение восстанавливается.

Методы MPEG (Pictures Experts Group) используют предсказывающее кодирование изображений (для сжатия данных о движущихся объектах вместе со звуком).

Так, если передавать только изменившиеся во времени пиксели изображения, то достигается сжатие в несколько десятков раз. Этот алгоритм сжатия используется также в стандарте H.261 ITU

Рисунок 21 – Алгоритм сжатия MPEG

Другой принцип воплощен в фрактальном кодировании (рисунок 22) при котором изображение представленное совокупностью линий описывается уравнениями этих линий.

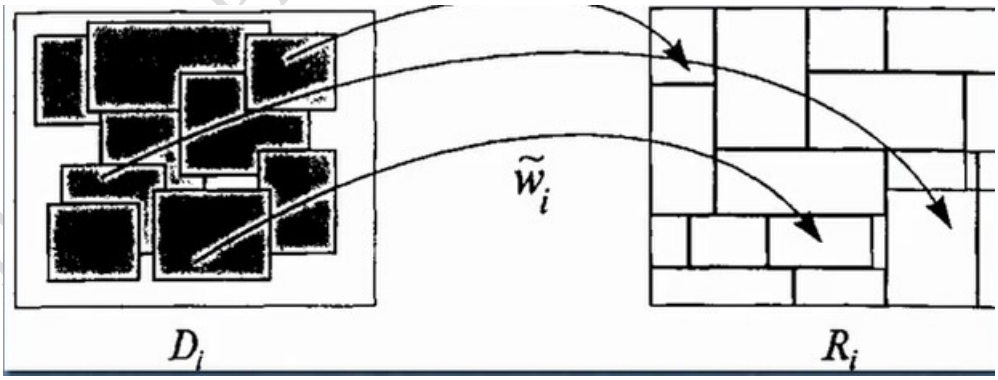


Рисунок 22 – Фрактальное кодирование

Боле универсален широко известный метод Хаффмана (рисунок 23) относящийся к статистическим методам сжатия. Идея метода – часто повторяющиеся символы нужно кодировать более короткими цепочками битов чем цепочки редких символов.

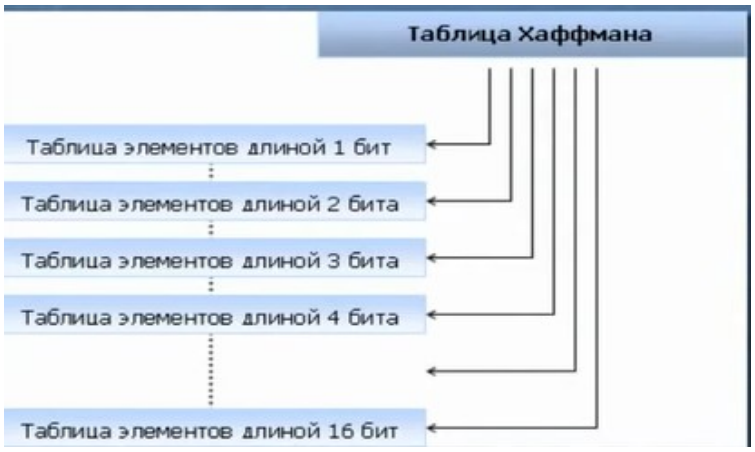


Рисунок 23 – Метод Хаффмана

Строится двоичное дерево, листья соответствуют кодируемым символам, код символа представляется последовательностью значений ребер (эти значения в двоичном дереве суть 1и 0), ведущих от корня к листу.

Листья символов с высокой вероятностью появления находятся ближе к корню, чем листья маловероятных символов.

Распознавание кода сжатого по методы Хаффмана, выполняется по алгоритму, аналогичному алгоритмам восходящего грамматического разбора.

Например, пуст набор из восьми символов (A,B,C,D,E,F,G,H) имеет следующие правила кодирования:

A::=10; B ::= 01; C ::= 111; D ::= 110;

E ::= 0001; F ::= 0000; G ::= 0011; H::= 0010

Тогда при распознавании входного потока 101100000110 в стек распознавателя заносится 1, но 1 не совпадает с правой частью ни одного из правил. Поэтому в стек добавляется следующий символ 0.

Полученная комбинация 10 распознается и заменяется на A. В стек поступает следующий символ 1, затем 1, затем 0. Сочетание 110 совпадает с правой частью правила для D. Теперь в стеке AD, заносятся следующие символы 0000 и т.д.

Недостаток метода заключается в необходимости знать вероятности символов. Если заранее они известны, то требуются два прохода: на одном в передатчике подсчитываются вероятности, на другом эти вероятности и сжатый поток символов передаются к приемнику.

Однако двух проходах сходность не всегда возможна. Этот недостаток устраняется в однопроходных алгоритмах адаптивного сжатия, в которых схема кодирования есть схема приспособления к текущим особенностям передаваемого потока символов.

Поскольку схема кодирования известна как кодеру, так и декодеру, сжатое сообщение будет восстановлено на приемном конце.

Обобщением этого способа является алгоритм, основанный на словаре сжатия данных. В ней происходит выделение и запоминание в словаре

повторяющихся цепочек символов, которые кодируются цепочками меньшей длины.

Интересен алгоритм “стопка книг”, в котором код символа равен его порядковому номеру в списке. Появление символа в кодируемом потоке вызывает его перемещение вначало списка

Очевидно,что часто встречающиеся символы будут тяготеть к малым номерам, а они кодируются более короткими цепочками 1 и 0.

Кроме упомянутых алгоритмов сжатия существует ряд других алгоритмов, например LZ – алгоритмы (алгоритмы Лемпеля-Зива). В частности, один из них (LZW) применен в протоколе V.42 bis.

Сжатие данных по методу Лемпеля-Зива

Метод использует следующую идею: если в тексте сообщения появляется последовательность из двух ранее уже встречавшихся символов, то эта последовательность объявляется новым символом, для нее назначается код, который при определенных условиях может быть значительно короче исходной последовательности (рисунок 24).

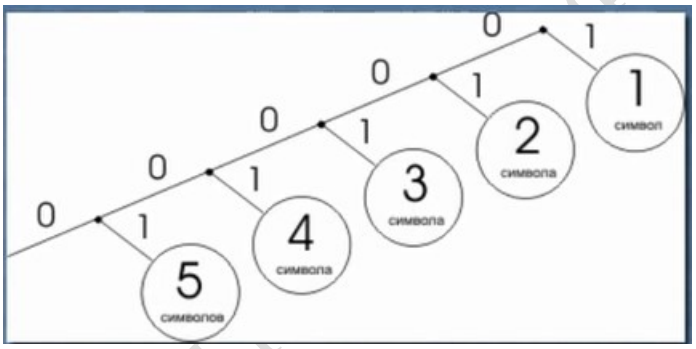


Рисунок 24 – Метод Лемпеля-Зива

В дальнейшем в сжатом сообщении вместо исходной последовательности записывается назначенный код. При декодировании повторяются аналогичные действия и потому становятся известными последовательности символов для каждого кода.

Одна из алгоритмических реализаций этой идеи включает следующие операции. Первоначально каждому символу алфавита присваивается определенный код (коды – порядковые номера, начиная с 0) при кодировании. Выбирается первый символ сообщения и заменяется на его код. Выбираются следующие два символа и заменяются своими кодами. Одновременно этой комбинации двух символов присваивается свой код. Обычно это номер, равный числу уже использованных кодов.

Так, если алфавит включает 8 символов, имеющих коды от 000 до 111, то первая двух – символьная комбинация получит код 1000, следующая – код 1001 и т.д. Выбираются из исходного текста очередные 2,3,…N. символов до тех пор, пока не образуется еще не встречавшаяся комбинация. Тогда этой комбинации присваивается очередной код, и поскольку совокупность А из первых N-1 символов уже встречалась, то она имеет свой код, который и

записывается вместо этих N-1 символов. Каждый акт введения нового кода назовем шагом кодирования. Процесс продолжается до исчерпания исходного текста.

При декодировании код первого символа, а затем второго и третьего заменяются на символы алфавита. При этом становится известным код комбинации второго и третьего символов. В следующей позиции могут быть только коды уже известных символов и их комбинаций. Процесс декодирования продолжается до исчерпания сжатого текста.

Сколько двоичных разрядов нужно выделять для кодирования? Ответ может быть следующим: число разрядов R на каждом шаге кодирования равно числу разрядов в наиболее длинном из использованных кодов (т.е. числу разрядов в последнем использованном порядковом номере).

Поэтому если последний использованный код (порядковый номер) равен 13=1101, то коды А всех комбинаций должны быть четырехразрядными при кодировании вплоть до появления номера 16, после чего все коды символов начинают рассматриваться как пятиразрядные (R = 5).

Например пусть исходный текст представляет собой двоичный код (рисунок 10), т.е. символами алфавита являются 0 и 1. Коды этих символов соответственно также 0 и 1.

Образующийся по методу Лемпеля-Зива код (LZ-код) показан во второй строке рисунка 25. В третей строке отмечены шаги кодирования, после которых происходит переход на представление кодов А увеличиным числом разрядов R.

Исходный текст	0.00.000. 01. 11. 111.1111. 110. 0000.00000. 1101. 1110.
LZ-код	0.00.100.001.0011.1011.1101.1010.00110.10010.10001.10110.
R	2 3 4
Вводимые коды	- 10 11 100 101 110 111 1000 1001 1010 1011 1100

Рисунок 25 – Сжатие данных по методу Лемпеля-Зива

Так, на первом шаге вводится код 10 для комбинации 00 и поэтому на следующих двух шагах R=2, после третьего шага R=3, после седьмого шага R=4, т.е. в общем случае R=K после шага 2^-1-1.

В приведенном примере LZ-код оказался даже длиннее исходного кода, так как обычно короткие тексты не дают эффекта сжатия. Эффект сжатия проявляется в достаточно длинных текстах и особенно заметен в графических файлах.

В другой известной реализации LZ-метода любая ранее встречавшаяся последовательность в сжатом тексте представляет собой совокупность данных:

- номер первого символа в ранее встречавшейся последовательности;
- число символов в последовательности;
- следующий символ втекущей позиции кодируемого текста.

Тема 5.4. ОБРАБОТКА ИНФОРМАЦИИ

Понятие обработки информации. Обработка данных и переработка информации. Основные виды обработки информации. Технологический процесс обработки данных. Технологическая сеть обработки данных. Типовые операции обработки данных. Сбор, регистрация, сортировка, поиск и выдача информации.

Технологический процесс обработки информации и его классификация

Внедрение и эффективное функционирование информационных технологий зависит от организации технологического процесса обработки информации на экономическом объекте.

Технологический процесс преобразует информацию, начиная с момента возникновения исходных данных и заканчивая получением ожидаемых результатов.

Построение технологического процесса обработки информации на предприятиях или в организациях определяется следующими факторами:

- особенностями обрабатываемой информации;
- типами решаемых задач;
- объемом обрабатываемой информации;
- требованиями к периодичности, срочности и точности обработки данных;
- соответствия временным регламентам взаимодействия производственных процессов и их элементов;
- типами, количеством и характеристикой технических средств обработки информации и т. д.

Эти факторы ложатся в основу организации технологического процесса, который включает перечень последовательности и способов выполнения операций, порядка работы специалистов и средств автоматизации, организацию рабочих мест, установление временных регламентов взаимодействия и т. д.

Организация технологического процесса должна обеспечить его экономичность, комплексность, надежность функционирования, высокое качество работ и т.д. Это достигается использованием системотехнического подхода к организации технологии обработки информации, которая строится на следующих принципах:

- интеграция обработки информации и возможность работы специалистов в условиях эксплуатации автоматизированных банков данных (АБД);
- распределенная обработка данных на базе развитых систем передачи информации;
- рациональное сочетание централизованного и децентрализованного управления посредством соответствующей организации технологического процесса обработки информации;
- моделирование и формализованное описание данных, операций их преобразования, функций и автоматизированных рабочих мест специалистов;
- учет конкретных особенностей экономического объекта, в котором реализуется информационная технология.

Технологические процессы обработки информации различаются составом и последовательностью операций, степенью их автоматизации, т.е. долей машинного и ручного труда, надежностью их выполнения и т.д. При

этом надежность обработки информации в технологическом процессе реализуется качеством выполнения основных операций и наличием разнообразных средств контроля.

В соответствии с этим выделяют большое количество технологических процессов обработки информации, которые можно классифицировать по различным признакам, приведенным в таблице 1.

Таблица 1. Классификация технологических процессов

Классификационный признак	Тип технологического процесса обработки информации
1. Тип организации технологического процесса	**Предметный тип** организации предполагает создание параллельно действующих технологических линий, специализирующийся на обработке информации и решении конкретных комплексов задач (учет труда и заработной платы, снабжение и сбыт и т.д.) и организующих пооперационную обработку данных внутри линии **Пооперационный (поточный) тип** построения технологического процесса предусматривает последовательное преобразование данных, согласно технологии, представленной в виде непрерывной последовательности сменяющих друг друга операций, выполняемых в автоматическом режиме
2. Степень централизации обработки данных	**Централизованный**, характеризующийся тем, что обработка информации и решение основных функциональных задач экономического объекта производится в центре обработки — центральном сервере, организованной на предприятии вычислительной сети, либо в отраслевом или территориальном информационно-вычислительном центре. **Децентрализованный**, основанный на локальном применении средств вычислительной техники, установленных на автоматизированных рабочих местах специалистов для решения конкретных функциональных задач. Децентрализованные технологические процессы не имеют централизованного автоматизированного банка данных, но обеспечивают пользователей средствами коммуникации для обмена данными между рабочими станциями сети. **Комбинированный**, характеризующийся интеграцией процессов решения функциональных задач на автоматизированных рабочих местах специалистов с использованием совместных баз данных и концентрацией всей информации экономического объекта в автоматизированных банках данных
3. Тип автоматизированного	Технологические процессы, выполняемые в системах обработки данных.

процесса управления	**Технологические процессы аналитической обработки данных** в системах подготовки принятия решений и экспертных системах. **Технологические процессы для разработки новых видов продукции** и получения чертежной и технологической документации в системах автоматизированного проектирования. Технологические процессы, выполняемые в системах электронного документооборота.
4. Отношение к ЭВМ	**Внемашинные** технологические процессы, имеющие подготовительный характер, т.к. их функционирование связано с получением исходных данных. **Внутримашинные**, связанные с хранением и обработкой полученной информации
5. Тип обрабатываемой информации	Технологические процессы обработки цифровых данных. Технологические процессы обработки текстовой информации. Технологические процессы обработки графической информации. Технологические процессы обработки мультимедийной информации. Технологические процессы на базе экспертных систем.
6. Тип используемого технического обеспечения	Технологические процессы обработки информации на базе персональных компьютеров. Технологические процессы обработки информации в локальных вычислительных сетях. Технологические процессы обработки информации в региональных сетях. Технологические процессы обработки информации в глобальных сетях
7. Режим обработки информации	**Пакетный** предусматривает выполнение обработки информации, оформленной в виде пакета заданий для ЭВМ под управлением ее операционной системы. Диалоговый предусматривает интерактивную связь пользователя с ЭВМ посредством устройств ввода информации (клавиатуры и др.), с которых возможен ввод команд, воздействующих на порядок работы программ обработки информации. **Режим разделения времени**, при котором компьютер используется несколькими пользователями одновременно, обычно при помощи системы квантования времени. Режим реального времени обеспечивает такую реакцию управления экономическим объектом, которая соответствует динамике его производственных процессов.

8. Тип информационного обеспечения	Технологические процессы, обрабатывающие локальные файлы. Технологические процессы, обрабатывающие локальные базы данных. Технологические процессы, обрабатывающие распределенные базы данных
9. Тип прикладного программного обеспечения	Технологические процессы, применяющие функционально ориентированные пакеты, используемые для автоматизации решения задач функциональных подсистем. Технологические процессы, использующие методом - ориентированные пакеты, применяемые для решения задач класса системы подготовки принятия решений. Технологические процессы на базе профессионально ориентированных пакетов, предназначенных для обработки различных типов данных.

В основу выбора типа технологического процесса обработки информации должна быть положена эффективность функционирования экономического объекта для достижения стратегических целей бизнеса.

В этом случае информационная технология в целом должна обеспечивать развитие бизнеса, его управляемость и качество, конкурентоспособность, снижение стоимости выполнения бизнес-процессов и т.д., а технологический процесс обработки информации должен поддерживать управленческую деятельность подразделений и структур предприятий и организаций, ложиться в основу управления информационными потоками экономического объекта.

Основным элементом технологического процесса является операция. Операции технологического процесса можно классифицировать по различным признакам, представленным на рисунке 26.

В соответствии с представленными классификационными признаками можно выделить следующие виды операций технологического процесса обработки информации.

1. Цель и место выполнения технологических операций. Выделяют четыре основных класса операций, которые отличаются, прежде всего, трудовыми и стоимостными затратами, связанными с их реализацией, целью и местом выполнения.

Первый класс включает операции по получению первичной информации, которая отражает состояние процессов в подразделениях промышленных предприятий, занятых производственной деятельностью. К данному классу операций относятся:

- сбор первичной информации, т. е. получение количественной характеристики показателей (например, количество изготовленных деталей, показания датчиков и счетчиков и т. д.);
- регистрация первичной информации, т. е. нанесение полученной информации на материальный носитель;
- передача первичной информации от места возникновения к месту обработки.

Рисунок 26 – Классификация операций технологического процесса обработки информации

Операции первого класса выполняются в основном на рабочих местах в производственных подразделениях вне места обработки информации. Данные операции являются самыми трудоемкими (до 50% трудовых затрат от трудоемкости всего технологического процесса обработки информации), дорогостоящими и дают наибольший процент ошибок в получаемых данных.

Второй класс включает операции ввода данных в ЭВМ. В процессе ввода возможна организация непосредственной передачи данных в вычислительную машину или перенесение первичной информации на промежуточные машинные носители, а затем занесение данных в ЭВМ. К этому классу задач относятся:

- прием, контроль и регистрация данных в пункте обработки информации;
- ввод данных в ЭВМ;
- контроль ошибок и загрузка данных в информационную базу;
- ведение информационной базы, включая такие операции, как корректировка информации, внесение дополнений и т. д.

Данный класс отличается достаточно высокой трудоемкостью (до 40% от трудоемкости всего процесса) и большим количеством допускаемых ошибок.

В современных информационных технологиях операции первого и второго классов совмещаются, когда в процессе сбора и регистрации первичной информации выполняется непосредственный ввод данных в ЭВМ.

Третий класс включает операции обработки данных в ЭВМ и получения результатной информации. Данный класс характеризуется наибольшей степенью автоматизации процессов, наименьшей трудоемкостью (5% от трудоемкости всех операций технологического процесса обработки информации) и наименьшим количеством допускаемых ошибок.

Четвертый класс операций ориентирован на обеспечение достоверности, своевременности получения и полноты результатной информации.

К основным операциям четвертого класса относятся:

- анализ и контроль полученных результатных данных;
- выявление и исправление ошибок по причине неправильности введенных исходных данных, сбоев в работе машины, ошибок пользователя, оператора или программиста.

Трудоемкость четвертого этапа составляет до 5% от трудоемкости всех процессов. Обычно этот класс операций выполняется при сложной аналитической обработке данных.

Тема 5.5. ХРАНЕНИЕ ИНФОРМАЦИИ

Физические основы хранения информации. Типы физических носителей. Специфика хранения информации. Интерфейсы физических носителей. Системы кодирования информации на физических носителях. Носители информации: их виды, история и будущее

Весь путь накопителей данных на примере романа Толстого — от наскальной живописи до молекулярных и кварцевых дисков.

Что ж, давайте посчитаем. А за единицу информации возьмём «Войну и мир» Толстого. Но для начала изучим немного матчасть.

Что такое носитель информации.

Под носителем информации понимают любой объект физического мира или структурную среду, которая умеет записывать, хранить, считывать и передавать данные.

Понятия «носитель» и «накопитель» обычно взаимозаменяемы, но иногда в информатике их разделяют. Так, носителем может быть что угодно: бумага, кассета, диск. А вот накопителем — только электронные устройства, которые умеют работать с информацией: классифицировать её, хранить, менять и перезаписывать.

Кстати, в английском языке тоже есть два понятия:

- Medium — это физические свойства материалов для хранения данных.
- Device differ или storage media — устройства для их чтения и записи.

Соответственно, носитель — это металлическая магнитная дисковая пластина, а запоминающее устройство — HDD.

Примеры носителей информации

Как мы уже говорили, любой объект, на который записали информацию и с которого её можно считать, — это носитель. Например:

- Бумажный лист с текстом, цифрами, иллюстрациями, графиками.
- Кассета на магнитной плёнкой с записью музыки.
- DVD-диск с фильмом.
- HDD-диск с операционной системой.

– Флешка с документами.

С некоторых носителей информацию можно считать напрямую — допустим, текст на бумаге мы воспринимаем визуально. Для чтения с других требуются специальные устройства — например, дисководы и электронные компоненты для дисков и HDD.

Виды электронных носителей информации

Существует несколько способов классификации информационных накопителей. Во-первых, их можно разделить по принципу записи:

Аналоговые — преобразуют непрерывный сигнал из внешнего мира (звук, изображение).

Цифровые — записывают входящую информацию в виде прерывного (дискретного) бинарного кода.

Например, фотоплёнка — это аналоговый носитель, а SSD-диск — цифровой.

Во-вторых, электронные носители информации делят на виды в зависимости от способов чтения и записи:

Магнитные: дискеты и HDD-диски.

Оптические: CD-ROM, DVD-ROM, Blu-ray.

Полупроводниковые: флеш-карты и SSD-диски.

Наибольшее распространение в XXI веке получили твердотельные накопители: жёсткие диски и флеш-карты. DVD и Blu-Ray всё ещё используют для дистрибуции фильмов, музыки, софта и видеоигр, но стриминг и онлайн-загрузка становятся всё более популярными.

Бумажные носители тоже никуда не делись — на них печатают книги, журналы, учебники, рекламные баннеры, брошюры и другую информацию, а вот грампластинки, перфокарты, VHS-кассеты устарели — ими интересуются только энтузиасты, коллекционеры и любители ретро.

Свойства носителей информации

Объект, который называют носителем информации, должен обладать следующими свойствами:

Долговечность. Записи не должны исчезать «в никуда» и умеют воспроизводиться по запросу пользователя.

Читаемость. Информацию можно получить напрямую или считав с помощью специального устройства.

Записываемость. Содержимое можно записывать или перезаписывать, менять структуру, порядок и свойства.

Точность и полнота. Данные на носителе сохраняются без ошибок и в полном объёме.

Надёжность. Данные защищены от воздействия внешней среды и их сохранность можно контролировать.

На протяжении всей истории человечества носители информации постоянно совершенствовались, но их основные свойства сохранились.

Древние носители информации

Начнём с времен, когда по земле ещё ходили неандертальцы. Даже неандертальцы умели накапливать и передавать информацию. Правда до алфавита они не додумались, а кодировали сообщения как попало — петроглифами.

Пещерные стены (40 тысяч лет до н. э.)

На стенах пещер, где жили древние люди, остались послания, которые дошли до нас спустя десятки тысяч лет — петроглифы. Вначале их наносили краской, но её смывало дождём, поэтому со временем стали применять гравировку или высекать изображения зубилом.

На древнейших петроглифах, которые датируются временами верхнего палеолита (около 40 000–20 000 лет до н. э.), изображали в основном животных. Некоторых из них уже нет — например, мамонтов и саблезубых тигров. Начиная с неолита (около 10 000 лет до н. э.), люди всё чаще стали рисовать себя, свои ритуалы и общинный быт.

О назначении петроглифов учёные спорят до сих пор. По одной из версий, они были делом тогдашних колдунов: познание мира происходило стихийно, науки не существовало, и люди могли верить, что рисование зверей на стенах сулит хорошую охоту.

Посмотрим, что бы получилось, если б среди них жил свой Лев Толстой, который внезапно придумал бы «Войну и мир». Если верить сайту «Ответ.Guru», бессмертный роман Льва Николаевича содержит 478 458 слов, или 2 521 613 знаков без пробелов, и занимает на диске 6,14 Мб.

Надо заметить, почерк у древних писателей был размашистый. Размеры самого крупного наскального изображения лебедя, обнаруженного на полуострове Кочковнаволок в Карелии, достигают 4,12 метров от хвоста до головы. Средний же размер петроглифа — 50 × 50 см (0,25 кв. м).

Иными словами, «Война и мир» в такой кодировке заняла бы как минимум 119 614,5 квадратных метров наскальной поверхности. А если с пробелами — раза в два больше выйдет, наверное. Это не считая многочисленных черновиков, с учётом которых полученную число ещё на пять умножить надо. В общем, картинка бы вышла размером примерно километр на километр.

Глиняные таблички (3500 лет до н. э.)

Высекать на камне тяжело — поэтому мягкая и податливая глина стала не только материалом для посуды, но и первым «листом». Древние шумеры начали использовать запись на глиняных табличках, придумали первые книги и библиотеки. Эту же технику использовали в Древнем Египте.

Люди лепили из глины плоские листы, а потом с помощью трёхгранной палочки кодировали данные клинообразными знаками. Затем таблицы сушили и ставили в определённом порядке в ящике — как листы в книге. Первая была титульной — на ней указывали автора и содержание. Ящики собирались в библиотеки — для этого на отдельных таблицах каталогизировать информацию. Чем не первая база данных или жёсткий диск?

Одна из крупнейших библиотек с клинописью была обнаружена во дворце правителя Ассирии — там хранилось более 30 тысяч глиняных таблиц

Хотя, если подумать, это тоже не самый удобный способ хранить информацию. Стандартный размер библиотечной таблички — 32 × 32 см, толщина — 2,5 см. Вмещались на ней, при самом убористом почерке клинописца, примерно 2300 знаков. Для «распечатки» подобным способом одного экземпляра «Войны и мира» потребовалось бы 1096 таких «кирпичиков».

Для сравнения, стандартный размер современного кирпича — 25 × 12 × 6,5 см. То есть одна глиняная табличка — это примерно 1,3 кирпича. Иными словами,

все четыре тома романа плюс эпилог в клинописной форме — где-то 1439 кирпичей.

Чтоб вы понимали: это шесть поддонов, не считая обложки. Больше пяти тонн. Хотите такую книжку дома у себя поставить? Заготовьте чуланчик размером 2,2 × 1,7 м. Это если в два ряда ставить и без обложки.

К слову, типовой кирпичный мини-завод смог бы производить не больше четырёх копий «Войны и мира» в день, без учёта работы клинописцев. Так себе книгопечатание. Возможно, именно поэтому и не была своего Льва Толстого в древней Месопотамии.

Папирус (3000 лет до н. э. — 1300 н. э.)

По оценкам учёных, папирус появился в Древнем Египте не раньше 5000 лет назад. Его изготавливали из одноименного растения, которое в изобилии росло на пресноводных болотах. Стебли папируса очищали от кожуры, нарезали на тонкие полоски, а потом накладывали друг на друга. Смесь высыхала на солнце и превращалась в лист, на который можно было наносить изображения и иероглифы. Папирус, как и хорошая бумага, делился на сорта — самые дорогие и качественные использовали при дворе, а дешёвые доставались купцам и мелким чиновникам.

Хрупкость папируса — миф. Низшие сорта действительно быстро приходили в негодность, но свитки из высококачественного материала могли сохраняться тысячи лет. Из-за своей дешевизны (лист папируса мог купить даже рабочий) и относительной простоты производства папирус использовали вплоть до XIII века н.э., пока его не вытеснила бумага.

Ну и наши традиционные расчёты. Один лист папируса на 2500 знаков без пробелов — это примерно 30 сантиметров тростникового стебля. «Война и мир» в самом экономном исполнении самым мелким почерком — это не менее тысячи папирусных листов, или не менее трёхсот срубленных тростинок, каждая по метру ростом.

Пергамент (500 лет до н. э. — XV век)

Пергамент — это обработанная кожа животных. Его начали использовать персы ещё в V веке до н. э. Это довольно грубый материал, более лёгкий и качественный вариант назывался «велень», но он появился только после XIII века.

До Средних веков бумага, папирус и пергамент были одинаково популярны. В Европе на папирусе писали религиозные книги, пока в начале XV века не изобрели книгопечатание. Пергамент отошёл на второй план, и его стала использовать в основном знать.

Вот этот способ уж точно не понравился бы известному вегетарианцу и зоозащитнику Льву Толстому. Один барашек — это от силы десять листов папируса. А на один экземпляр «Войны и мира», как мы уже знаем, таких листов требуется не менее тысячи. В общем, один роман стоил бы как минимум сотни загубленных ягнят.

Пальмовые листья (500 лет до н. э. — XVIII век)

Пальмовые листья использовали в Индии и Юго-Восточной Азии. Текст и изображения наносили пером с чернилами, либо делая надрезы. Листья собирались в стопку и связывались нитью, образуя книгу. Их покрывали воском, чтобы не сгнили, и лемонграссом для отпугивания насекомых, которые могли эти листья съесть.

Пальма росла в изобилии, а техника производства была простой, поэтому на пальмовом листе писали вплоть до XVIII века. Не будем утомлять вас расчётами: они, в принципе, идентичны приведённым выше расчётам с папирусом. Все пальмы Вьетнама ободрать надо было бы, чтоб каждому советскому школьнику по экземпляру «Войны и мира» сделать.

Береста (IX–XV века)

Береста — это верхний слой берёзы, который использовали как бумагу. На поверхности царапали слова грифельной или заточенной деревянной палочкой, а потом протирали её углем или сажей. Берестяные грамоты в Древней Руси использовали с IX по XV век.

Больше всего берестяных табличек обнаружено в местах раскопок Великого Новгорода. В основном послания на бересте использовали как быстрый способ передачи информации между городами: в них были бытовые указания, требования вернуть долг или торговые предложения. Быстрый-то он, конечно, быстрый. Но по расходу материала — это даже по сравнению с папирусом растратно. Возьмём, к примеру, стандартную берёзку высотой три метра и диаметром ствола 20 см. Вспомним школьную формулу расчёта площади цилиндра и вычислим, сколько на ней бересты:

$S = h * \pi * D = 300 * \pi * 20 = 18\,840$ кв. см,

или примерно 30 привычных нам листов А4.

Даже если бы нашёлся уникум, которым смог бы нацарапать грифелем по бересте столь же убористо, как мы авторучкой по бумаге, — это от силы 1500 знаков на один берестяной лист А4.

Итого: Одна «Война и мир» = 2 521 613 / (1500 * 30) = 56 берёзок.

Бумага (200 лет до н. э. — по настоящее время)

В древнем Китае бумагу делали из бамбука, волокон конопли и шёлка. С первого века начали смешивать в массу тутовое дерево, ткань, волокна и золу, образуя из неё листы.

Технология производства бумаги постоянно совершенствовалась и распространялась по миру: сначала в Японии и Корее, затем в Европе. В XV веке было изобретено книгопечатание, и бумага вытеснила все остальные носители информации: бересту, папирус, пергамент.

Промышленное производство бумаги из древесной целлюлозы появилось в XVIII веке. Книги, газеты и журналы стали доминирующими носителями данных. Скорость распространения информации кратно выросла, библиотеки стали «храмами знаний» — всё это повлияло на научно-технический прогресс.

Действительно, компактность хранения информации возросла кратно. «Война и мир» в стандартном 14-томном собрании сочинений Л. Н. Толстого 1951 года (тома с 4-го по 7-й) занимает такой объём:

- том 4 — 361 страница;
- том 5 — 377 страниц;
- том 6 — 406 страниц;
- том 7 — 363 страницы.

Да и тираж: 300 000 экземпляров. Просто, массово, надёжно. По этой, в частности, причине роман и ввели в школьную программу, навсегда привив миллионам учеников «любовь» к русской классике.

Носители информации в XX веке

В конце XIX и начале XX века индустриализация привела к повсеместному внедрению электричества. Появились новые виды транспорта: пароходы, автомобили и поезда, а также более быстрые средства связи — телефон и телеграф. Бумага всё ещё оставалась основным носителем информации, но учёные и предприниматели начали разрабатывать новые устройства, которые постепенно изменили мир.

Перфокарты (начало XIX века — середина XX века)

Перфокарты — один из первых накопителей данных, который можно было декодировать только с помощью машины. Расположенные на них в определённом порядке отверстия были, по сути, прототипом двоичной системы.

Перфокарты использовали при вышивании узора на ткацких станках, для вычислений и классификации информации. Их настоящим триумфом стала перепись населения в 1890 году. Для обработки результатов прошлых переписей требовалось не менее десяти лет. Внедрение перфокарт сократило этот срок до трёх месяцев.

В модернизированном варианте — как перфоленты — перфокарты дожили аж до восьмидесятых годов прошлого века. Их активно применяли в тогдашних ЭВМ — например, в первом программируемом компьютере «Марк I», а затем на заводах для станков с ЧПУ.

Скорость записи составляла максимум 150 байт в секунду, чтения — 1,5 килобайта в секунду. Немного, но вполне достаточно, чтобы обрабатывать математические операции. Ну, или чтобы «Войну и мир» прочитать чуть больше, чем за час. Блестящий результат, если учесть, что большинству людей для этого целой жизни не хватает.

Правда, бумажная лента была хрупкой, и её нельзя было редактировать — систему хотели улучшить, заменив бумагу на пластик, но технология была слишком дорогой. Появившиеся позже магнитные ленты предлагали более удобный способ записи и чтения.

Восковый валик (1857–1920-е годы)

Эпоха нового подхода к накопителям началась с записи звука. В конце XIX века появился первый фонограф. В отличие от музыкальных шкатулок, он умел не только воспроизводить звук, но и записывать его: на цилиндрическом восковом носителе игла преобразовывала звук в борозды, которые потом можно было считать обратно. Максимальное время записи составляло около двух минут.

Изобретатель устройства Томас Эдисон составил целый список сценариев использования устройства: диктовка писем, обучение языкам и даже запись телефонных звонков. Была также провальная идея использовать фонограф в больших куклах, но подобные игрушки пугали детей.

Фонограф Эдисона, 1899 год

В начале XX века вместо воскового цилиндра в фонографах начали использовать первые диски из мягких металлов, на которые можно было записывать до четырёх минут звука. Со временем технология улучшилась и фонограф превратился в электрический граммофон с грампластинками.

Грампластинка (1910-е годы — наше время)

Диски из цинка и эбонита существовали параллельно с восковыми, но с их помощью нельзя было перезаписывать информацию. Зато у них было другие

технические преимущества — форма, которую легко штамповать и тиражировать. В начале XX века появился более удобный винил, а в конце 1920-х — электроакустическая запись через микрофон: диапазон расширился, качество записи улучшилось.

Устройство для записи медных пластинок. Германия, 1930-е годы В 1930-х технологии ещё не позволяли плотно записывать информацию, поэтому на одну пластинку помещалась одна песня: альбомы продавались в коробках. После Второй мировой войны появились «долгоиграющие пластинки», в которые уже помещались целые альбомы.

До появления аудиокассет и компакт-дисков грампластинки оставались основными накопителями для воспроизведения музыки. Впрочем, некоторые аудиофилы предпочитают их до сих пор. А что? Опера Прокофьева «Война и мир» — это всего четыре диска. А ведь там к словам ещё и музыку добавили! По сравнению с тоннами глиняных табличек — просто гигантский прогресс, согласитесь.

Фото- и киноплёнка (1850–1950-е годы)

Фотоплёнка и киноплёнка — носители визуальной информации, которые хранят в себе «микрооттиски». Плёнка состоит из нескольких слоев, один из которых светочувствительный. Впервые фотоплёнки появились в конце XIX века, в кинокамерах их стали массово использовать в начале XX века. Оба носителя выпускают до сих пор, пусть и не в таких объёмах, как раньше.

Чёрно-белая киноплёнка

Интересно применение фотопленки для архивации печатных материалов — микрофиши. Наверняка вы не раз видели в фильмах, как главные герои в поисках информации отправляются в библиотеку и на больших машинах листают старые газеты.

Данные, которые выводят на экран, — это копии, записанные в формате микрофиши. На одну микрофишу размером 7,5 на 12 см могло уместиться до 130 страниц книжного текста. До массового распространения компьютеров такая архивация позволяла существенно освободить полки от бумаги.

Кстати, именно киноплёнка позволила советским детям изобрести один из главных школьных лайфхаков: зачем целое лето на чтение «Войны и мира» тратить, если можно за семь часов кино Бондарчука посмотреть, все четыре серии?

Магнитная лента и ленточные накопители (1930–1990-е годы)

Магнитнитную ленту изобрели в Германии в 1927 году — первоначально на тонкую бумагу наносили напыление порошком оксида железа. В 1932 году компания AEG представила первое коммерческое устройство для чтения и записи — Magnetophon K1.

Технология совершенствовалась — в 1950-х на магнитную ленту стали записывать видео, а в 1970-х появился простой и надёжный формат VHS, который популяризировал видеокассеты как основной носитель для фильмов и телепередач. Параллельно с этим компакт-лента завоёвывала аудиорынок, вытесняя винил.

Использовать магнитную ленту как хранилище данных в компьютерах стали в 1951 году — первые ленточные накопители могли хранить не более нескольких килобайт данных.

К 1970-м появился стандарт 9-дорожной ленты, который мог вместить до 140 мегабайт, а в 1990-х — технология записи DLT с потрясающей ёмкостью в 800

гигабайт. 133 320 копий «Войны и мира» разместить можно, если что. И ещё место останется!

Ленточные накопители были надёжными, быстрыми, потребляли минимум энергии, но скорость доступа к произвольным участкам была медленной — ленту нужно было отматывать к определённому месту.

В современном мире ленточные накопители никуда не делись. Сейчас существует два популярных формата — LTO и IBM 3592, которые используют в ленточных библиотеках. Такие базы данных дешевле и энергоэффективней дисковых. Библиотеки могут хранить тысячи магнитных лент по несколько гигабайт каждая — по запросу робот быстро находит и производит чтение-запись с ленты.

Гибкий диск (1960-е — 1990-е)

Гибкий диск (флоппи-диск) — это круглый пластиковый носитель с магнитным покрытием, заключённый в пластиковый контейнер. Для чтения нужен специальный дисковод. Дискеты стали массово использовать в начале 1970-х в компьютерах IBM, но сейчас их считают устаревшими носителями. Но память о них осталась в софтовых иконках сохранения данных: многие из них отрисованы как флоппи-диск.

Технологий записи было несколько. В ранних версиях использовали FM-кодирование — оно записывало два тактных сигнала на один бит данных. Усовершенствованная технология MFM объединила два такта в один сигнал, что увеличило плотность записи в два раза. Затем в конце 1970-х появилась M2FM с дополнительными тактами.

У дискет существовало несколько форм-факторов: 8, 5,25 и 3,5 дюймов в диаметре. Последние назывались микрофлоппи-дисками и были распространены в конце 1980-х и начале 1990-х. Объём был небольшой — от 1,44 до 2,88 мегабайта (то есть чуть меньше половины «Войны и мира»), но этого вполне хватало, чтобы хранить документы, софт, видеоигры и даже операционные системы.

Объёмы файлов росли, а использование десятка флоппи-дисков было неудобным. Последняя попытка улучшить гибкий диск — SuperDisc в конце XX века — не увенчалась успехом: он мог вместить до 140 мегабайт данных, но проиграл «войну форматов» компакт-дискам, DVD и онлайн-дистрибуции.

Современные носители информации

Перфокарты, ленточные накопители и гибкие диски в какой-то момент перестали соответствовать возросшим требованиям. Компьютерам нужны были более быстрые и компактные способы чтения и записи.

Компакт-диски. На компакт-дисках (CD) был принципиально новый способ чтения и записи информации — с помощью оптического лазера. На диск из поликарбоната наносили специальный слой из металла, который хранит данные в микровыемках, а луч лазера отражается от этого слоя и считывает данные.

Технология впервые была использована для музыкальных записей в 1970-х, но уже в конце 1980-х была адаптирована для компьютеров (CD-ROM). Затем появились диски для однократной (CD-R) и многократной записи (CD-RW).

Объём данных на CD-ROM не превышал 700 мегабайт, но затем в 1996 году появился DVD с более плотной структурой слоя. За счёт нового лазера с меньшей длиной волны на него удаётся записывать до 17 гигабайт данных.

В 2006 году начали появляться Blu-ray Discs — для этих накопителей использовали ещё более коротковолновой лазер синего цвета. Плотность записи составляла до 50 гигабайт, что было особенно важно при возросшем качестве кинофильмов и увеличившихся объёмах софта. Появление дисков совпало с развитием интернета — популярность стримингов и онлайн-дистрибуции привела к тому, что диски для чтения и записи используют сейчас редко, а большинство моделей ноутбуков идут без встроенного дисковода.

Жёсткий диск (HDD) — устройство, в котором используются жёсткие пластины из алюминия или стекла покрытые специальным магнитным составом. Они заключены в металлический контейнер с блоком электроники, а сами диски находятся в герметичной зоне. Внутри нет вакуума, но часто закачан чистый воздух, чтобы избежать появления пыли.

Для чтения записи используется целая система устройства: электромотор вращает диски со скоростью от 7200 оборотов в минуту, а считывающая головка (коромысло) вводит или декодирует информацию.

Данные записываются не произвольно, а в строгом порядке по секторам. Размер одного сектора — 512 байт. Группа секторов образует кластеры, и как раз между ними и происходит обмен информацией.

Существует несколько методов записи:

Продольный (CMR). Биты записываются головкой над поверхностью дисков: так намагничивают сотни миллионов «доменов» — дискретных областей. Плотность записи составляет около 20 гигабайт на квадратный сантиметр, с 2010-х такая технология практически не используется.

Перпендикулярный (PRM). Биты сохраняются в горизонтальных доменах: плотность записи выше — от 60 гигабайт на квадратный сантиметр, до, теоретически, до терабайта на дюйм. Наиболее популярный сейчас метод.

Черепичный (SMR). Дорожки на диски «накладывают» друг на друга как в черепичных крышах, чтобы увеличить плотность чтения и записи головкой. У технологии есть минус — низкая скорость записи и перезаписи данных. Используется редко.

Когда жёсткие диски появились, то стояли они невероятно дорого — десятки тысяч долларов и вмещали в себя мегабайты данных. Неудивительно, что более практичные гибкие диски использовали для софта и операционных систем вплоть до 1980-х — к этому времени технологии HDD стали совершеннее и доступнее.

Твердотельные накопители (SSD) и flash-карты вместо магнитных дисков накапливают информацию на микросхемах, что в сотни раз повышает скорость записи и чтения. Впервые технология появилась в конце 1970-х годов и использовалась для суперкомпьютеров. С 2010-х появились первые доступные SSD-накопители с объёмом памяти 128 гигабайт, которые стали применять в компьютерах и ноутбуках.

Твердотельные накопители делятся на несколько групп:

Внешние flash-карты. Карты памяти (SD, microSD) и USB-флеш-карты («флешки»). Благодаря универсальному стандарту записи и интерфейсам, данными можно свободно оперировать и переносить с устройства на устройство — альтернатива сценарию флоппи-диска.

Встраиваемые SSD-диски. Альтернатива HDD в качестве запоминающего носителя на устройствах. Первые SSD-диски повторяли разъём подключения

SATA, современные твердотельные накопители подключаются через более быстрый PCI Express.

В большинстве своём твердотельные накопители работают на архитектуре NAND SSD, что обеспечивает высокую ёмкость, скорость и низкое потребление электричества. Есть ещё энергозависимые — RAM SSD, они работают по принципу оперативной памяти и ещё быстрее, но стоят заметно дороже первых.

У SSD много преимуществ по сравнению с HDD: малые габариты, отсутствие движущихся частей, низкая чувствительность к внешним электромагнитным полям, более надёжное сохранение данных.

Но есть и недостаток — ограниченное количество циклов для записи: у недорогих «флешек» он может составлять менее тысячи циклов, на внутренних SSD — до десятка тысяч. В любом случае, срок службы твердотельных накопителей составляет около пяти лет — после этого будут всё чаще появляется ошибки. Всё дело в ячейках памяти — они пока не могут выдерживать долгих циклов записи.

Накопители будущего

Объёмы информации продолжают расти, а кремний для чипов становится всё более дефицитным товаром. Учёные ищут перспективные направления, которые помогут записывать данные альтернативным способом. Вот некоторые примеры технологий, которые могут появится в ближайшем будущем.

ДНК

Биолог из Гарварда Джордж Чёрч в 2017 году продемонстрировал, что ДНК годится не только для передачи генетического кода белковых организмов. Информацию можно сохранять прямо в геном, а потом считать обратно. В эксперименте изображение и GIF-анимацию из шести кадров перекодировали в цепочку последовательности нуклеотидов, а затем синтезировали в искусственные ДНК.

Затем их внесли в кишечные палочки, и она сохранила данные по системе CRISPR — эту технологию используют живые организмы, чтобы управлять иммунитетом. Клетки были культивированы и разрослись до большой колонии, а затем учёные выделили из них ДНК и реконструировали данные.

Эту технологию пока сложно масштабировать для коммерческого производства (да и данные сохраняются не на 100%), но она уже доказывает, что в теории накопителями могут быть сами гены.

Колонии бактерий

В 2021 году учёные из Колумбийского университета превратили колонию бактерий в запоминающее устройство и подключили её к компьютеру для чтения и записи информации. Суть метода описана в статье журнала Nature — учёными была разработана электрогенетическая основа для прямого хранения цифровых данных в живых клетках. С помощью электрических сигналов они смогли закодировать двоичные данные в CRISPR-массивах бактериальных клеток.

Пока удалось записать только 72 бита информации, но зато эти данные могут храниться в естественной среде на протяжении многих поколений бактерий. Это открывает перспективные способы обмена информацией между кремниевыми и углеродными веществами.

Синтетический пластик

В 2021 году учёные из Техаса использовали для записи молекулы синтетического пластика. В его состав входят аминоспирты — в них удалось зашифровать буквы английского алфавита. Потом из этой конструкции построили макромолекулу, содержащую строку из книги Джейн Остин «Мэнсфилд-парк», и при расшифровке удалось восстановить 98% данных.

Молекулярный диск

Группы молекул тоже могут накапливать информацию. Учёные из университета Брауна смогли записать и считать небольшое изображение размером в пару килобайт с помощью молекулярного жёсткого диска.

Как это работает: информация записывается в искусственный раствор «метаболома» (так биологии называют набор молекул для регулирования процессов метаболизма в организме). Внутри — группы малых органических молекул-метаболитов.

Дальше дело техники — берём бинарную систему, где можно зашифровать что угодно, и воплощаем её: отсутствие или наличие метаболита означает, соответственно, ноль и единицу.

Робот помещает в тысячи капель размером в один нанолитр (одна миллиардная литра) на металлические пластины, и они выстраиваются в связи. Считывание с такого диска происходит с помощью химического анализа — высушенную металлическую пластину анализируют масс-спектрометром и расшифровывают обратно.

Технология не идеальна — метаболиты могут неконтролируемо воздействовать друг на друга и портить данные, а для накопителя крайне важно корректно хранить информацию. С другой стороны этот недостаток может стать фишкой — молекулярные жёсткие диски смогут не только хранить данные, но и манипулировать ими как компьютер.

Кварцевый носитель

В 2018 году Илон Маск вывел на орбиту Земли Tesla Roadster. В электромобиле находился диск с трилогией фантастических романов Айзека Азимова «Основание» про далёкое космическое будущее человечества. Этот диск Маску ранее подарил фонд сохранения человеческих знаний Arch Mission Foundation, но интересен не символизм, а технология записи данных — книга была записана на кварцевом стекле с помощью фемтосекундного лазера (оптический квантовый генератор, который посылает импульсы лазерного излучения). Эта технология известна как «кварцевый диск», Superman memory crystal, или Eternal 5D.

Технология записи на кварцевый носитель была разработана ещё в 1993 году, но в реальность её удалось воплотить только через 20 лет. В 2013 году учёные из Саутгемптонского университета смогли записать 300 килобайт, выжигая лазером на кристалле микроскопические точки. Метод Eternal 5D — это запись точек слоями на расстоянии пары миллионных метра друг от друга в пяти измерениях: длина, ширина, высота, ориентация и размер. Похоже на запись CD, только более прогрессивными способом.

Выжженные точки меняют характеристики кристалла и поляризацию света. Для считывания информации достаточно пропустить через него луч и считать данные с помощью поляризатора и микроскопа.

Внесение информации на кварцевый диск с помощью фемтосекундного лазера

Технология перспективная, но всё ещё дорогая — фемтосекундный лазер стоит десятки тысяч долларов и имеет большие габариты. Но зато кварцевые носители в перспективе более вместительны, чем жёсткие диски и SSD: на кварцевое стекло можно вместить более 300 терабайт данных, а информация при комнатной температуре сохраняется миллиарды лет.

Сам же кварцевый носитель может выдержать до 1000 градусов по Цельсию. Неудивительно, что диск с «Основанием» был помещён в Tesla Roadster — даже в открытом космосе данные на нём не пострадали.

Сохранение информации для будущих поколений

Носители и накопители информации быстро устаревают. В 2008 году NASA потребовалась информация о свойствах лунной пыли, но все данные об экспедициях на Луну хранилась на магнитных лентах… В итоге рабочее устройство для чтения удалось найти только в музее.

Сейчас информацию передают и хранят в основном с помощью удалённых хранилищ. Кино, музыка, видеоигры, программы и сервисы доступны через стриминг и онлайн-сервисы. Облачные технологии помогают рассредоточить информацию по интернету, сделать её чтение доступным с любого устройства, но что делать, если связи не будет или она вообще пропадёт на долгие годы, а то и столетия?

Угроза общепланетарной катастрофы ставит перед учёными нетривиальную задачу организовать цифровой архив, который поможет сохранить данные. Силами фондов ведутся разработки защищенных архивов. Например, уже работает «Арктический мировой архив» (Arctic World Archive, AWA), который расположен на Шпицбергене.

Фонд Вислявы Шимборской передаёт архив работ польской поэтессы и лауреата Нобелевской премии (слева). Делегация из Казахстана держит флаг на фоне входа в архив после передачи данных о конституции, флаге, гербе и гимне (справа)

Данные в AWA хранят по технологии усовершенствованной микрофиши. Пленку пакуют в пакеты и помещают в стальные контейнеры. Например, там есть исходный код GitHub, NFT-токены, цифровое искусство, труды лауреатов Нобелевской премии и даже конституция Казахстана. Проектировщики планируют, что данные могут храниться в Арктике несколько тысяч лет.

Другой проект — «Память человечества» (Memory of Mankind, MOM), решил вообще отказаться от цифровых носителей и использует в качестве накопителей информации керамические таблицы. Технология была навеяна шумерскими глиняными табличками. Цель архива — создать «капсулу времени» и построить образ эпохи для потомков. Располагается архив в соляной шахте в Австрии.

Группа учёных из Гарвардского университета решила пойти дальше и предлагает сделать бэкап на Луне. В своей работе A Lunar Backup Record of Humanity они предлагают не только спрятать архив в лунной пыли, но и держать с ним удалённую связь. Предложение фантастическое, но вполне реализуемое — с архивом можно будет связаться с помощью лазера. Уже сейчас можно передавать около 1015 байт в год — хватит для накопления данных о литературе, научных

открытиях и генетической информации, которая пригодится будущим поколениям. Такой низкий битрейт связан с мерцанием, атмосферными явлениями и постоянным рассеиванием луча.

Миссия Orion продлится около десяти дней: пилотируемый корабль с Земли сделает круг вокруг Луны и вернётся обратно

Систему связи через лазер можно доработать: если создать комплекс из лазерных передатчиков, то реально получить постоянный поток данных в 622 Мбит/с. Такие возможности — не технологии далёкого будущего: у NASA есть подобная система оптической связи Orion Artemis II (O2O). В 2023 году её испытают на космических кораблях Orion — такой быстрый канал связи обеспечивает не только полное пилотирование с Земли, но и трансляцию в 4K-качестве.

Накопители информации прошли большой путь от наскальных рисунков до записи лазером и архивации на Луне. Будущее накопителей информации — это развитие как цифровых, так и биологических технологии, который помогут не только хранить возросшие объёмы данных, но и взаимодействовать с ними на принципиально новых уровнях.

Тема 5.6. ТЕХНОЛОГИИ ХРАНЕНИЯ ИНФОРМАЦИИ

Основные аспекты хранения информации. Базы данных. Хранилища данных и их классификация. Технология OLAP. Основные сферы применения технологии OLAP-кубов данных. Хранилища данных. ЦхОДы и их классификация.

Все ЦхОДы можно классифицировать по их предназначению:

- Интернет-ЦхОД
- ЦхОД предприятий и организаций
- ЦхОД публичных провайдеров.

ЦОД также можно подразделить по способу реализации на стационарные, модульные и контейнерные (разновидность последних двух – т.н. «микро-ЦОД»). Причем в состав стационарных ЦОДов в качестве составных частей могут входить модульные и контейнерные, что удобно для целей резервирования, либо в случае необходимости быстрого расширения.

Разновидностью модульных ЦОДов являются т.н. «микро-ЦОД», обычно представляющие собой одну стойку, в которой смонтированы все компоненты ЦОД: серверы, системы хранения, коммутаторы и системы охлаждения и вентиляции. Такие ЦОД могут быть использованы в небольших предприятиях, а также в подразделениях средних и больших предприятий, для автономного хранения и обработки внутренней информации.

Стационарные ЦОД

Внешний вид типового здания для размещения стационарного ЦОД показан на рис. 4. На крыше видны теплообменники климатической установки.

Рисунок - Внешний вид стационарного ЦОД

Слева расположен главный вход, спереди – грузовые подъезды.

Примерное внутренне расположение оборудования и инфраструктуры в таком ЦОД, состоящего из двух надземных и одного подземного этажей, показано на рис. 5.

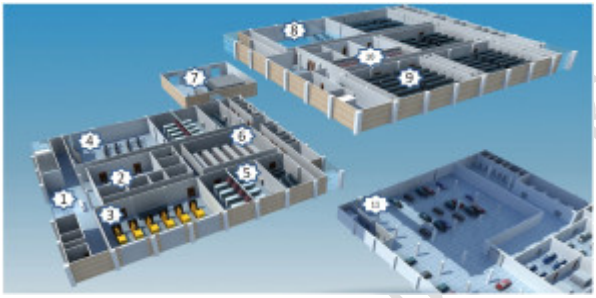


Рисунок 5 - Внутреннее расположение оборудования и инфраструктуры ЦОД

На рисунке цифрами обозначены:

1. Главный вход.
2. Служебные помещения.
3. Дизель-генераторы для питания ЦОД при отсутствии внешнего электроснабжения.
4. Центр управления ЦОД и его удаленных филиалов.
5. Шкафы электропитания.
6. Аккумуляторные батареи
7. Дополнительное помещение для расширения мощности и емкости ЦОД (могут использоваться модульные или контейнерные блоки).
8. Помещение для прикладных систем, например, системы «Безопасный город».
9. Помещения для оборудования ЦОД: серверов, систем хранения и сетевого оборудования.
10. Противопожарная система
11. Вспомогательные помещения, автостоянка и пр.

Модульные ЦОД

Модульный ЦОД представляет собой законченную конструкцию, включающую в себя все элементы ЦОД: серверы, накопители и коммутаторы, а также инфраструктурные элементы систем микроклимата и пожаротушения.

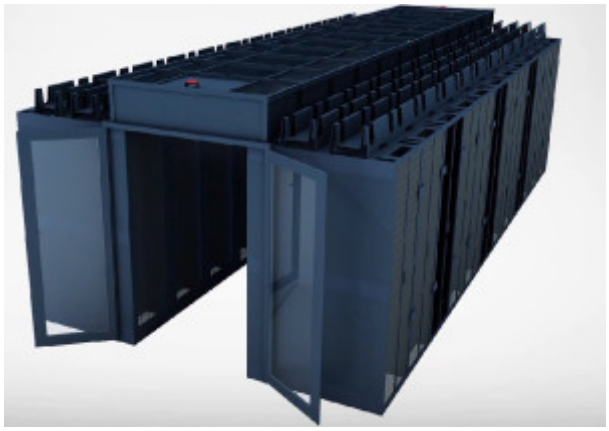


Рисунок 6 - Внешний вид модульного ЦОД

Как правило, модульный ЦОД имеет отдельный вход, герметизирующий его внутренности от среды помещения, в котором он устанавливается. Хотя основное предназначение модульного ЦОД – использование его в неприспособленных помещениях, его также можно использовать в зданиях стационарного ЦОД.

Контейнерные ЦОД

«Контейнерные ЦОД» (или «мобильные ЦОД») так называют потому, что их принято размещать в стандартных грузовых «сорокафутовых» контейнерах. Это делается с целью повышения оперативности развертывания, снижения затрат на создание ЦОД.

Контейнерные ЦОД могут быть весьма эффективны при различных экстренных ситуациях, например, создания оперативных штабов по ликвидации последствий стихийных бедствий, а также для задач оперативной деятельности малого и среднего бизнеса, где строительство больших стационарных ЦОД экономически невыгодно.

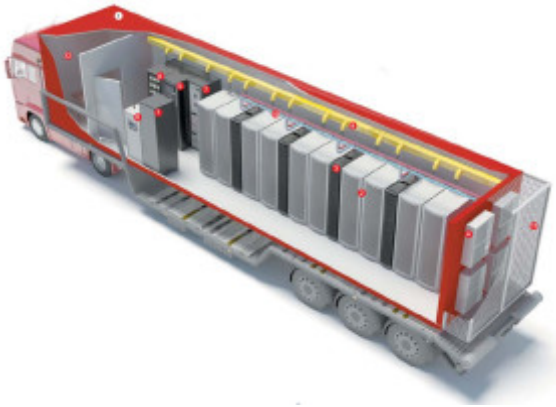
Контейнерные ЦОД можно также использовать для оперативного расширения емкости больших стационарных ЦОД, а также для целей резервирования, чтобы повысить уровень ЦОД в случае необходимости.

Внешний вид такого контейнера для размещения оборудования ЦОД оказан на рис. 7.

Рисунок 7 - Внешний вид контейнера ЦОД

Такие контейнеры удобно перевозить автотранспортом или другими видами грузового транспорта (морским, железнодорожным) с использованием стандартных погрузочно-разгрузочных средств.

Контейнеры могут быть как полностью интегрированные (рис. 8), так и модульные (рис. 9).



1. Контейнер размерами 12200x2500x2700 мм
2. Серверная стойка
3. Встроенные вентиляторы
4. Внешняя установка кондиционирования
5. Модульный источник бесперебойного питания
6. Батарея
7. Распределительные шкафы электропитания
8. Установка пожаротушения
9. Кабельные каналы
10. Трубы для циркуляции жидкости системы охлаждения
11. Защитная решетка для наружной установки
12. Комната администратора

Рисунок 8 - Интегрированный контейнерный ЦОД при транспортировке

Оборудование ЦОД в контейнерах может располагаться как вдоль (рис. 2-10), так и поперек контейнера (рис. 9).

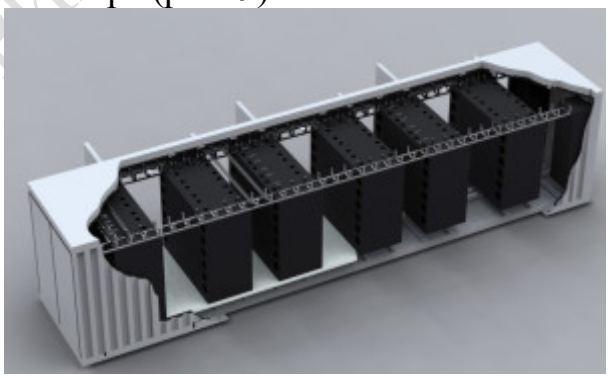


Рисунок 9 - Различные варианты расположения оборудования ЦОД в контейнере

При поперечном расположении оборудования, по сторонам контейнера обычно устраиваются двери, и контейнеры монтируются длинными сторонами друг к другу, таким образом, образуются проходы вдоль рядов оборудования внутри нескольких контейнеров, образующих единый модуль.

Рисунок 10 - Пример расположения оборудования ЦОД в 40-футовом контейнере и несколько контейнеров в едином модуле

Контейнерные ЦОД могут быть быстро доставлены и развернуты непосредственно на месте монтажа. Обычно, контейнеры с оборудованием монтируют под легковозводимым укрытием, хотя конструкция контейнера способна обеспечивать работу на открытом воздухе (outdoor) в довольно широком температурном диапазоне и различных погодных условиях.

Рисунок 11 - Доставка контейнеров на место монтажа

Контейнеры можно также устанавливать в два этажа, тем самым достигая экономии требуемой площади.

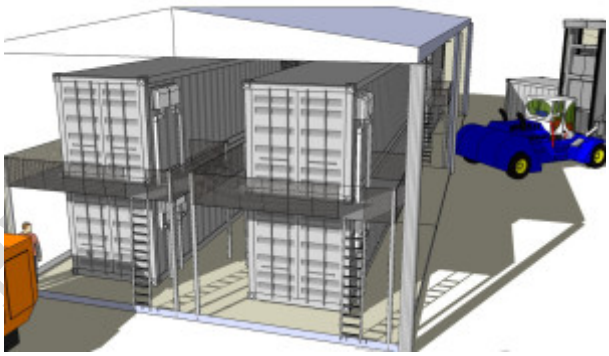


Рисунок 12 - Установка контейнеров в два этажа. (Изображение: Huawei Technologies)

Микро-ЦОД

Для небольших предприятий и филиалов крупных и средних предприятий могут подойти ЦОД небольшого размера («микро-ЦОД»), представляющее собой т.н. решение «все в одном» (All-In-One).

Рисунок 13 - Пример размещения микро-ЦОД в офисе

Микро-ЦОД могут быть как комнатного, так наружного исполнения. В последнем случае элементы ЦОД помещаются в термозащищенный вандалоустойчивый шкаф.

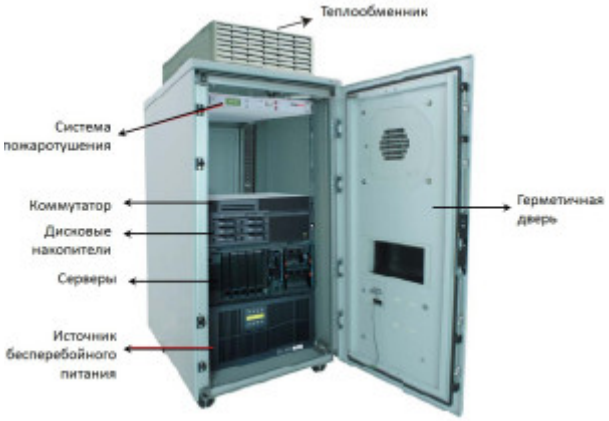


Рисунок 14 - Пример Микро-ЦОД наружного исполнения (Изображение: Canovate)

Строго говоря, микро-ЦОД не относятся к аппаратным решениям систем СС, ибо такие решения в основном, применяются для внутренних корпоративных систем предприятий.

Рисунок 15 - Пример размещения микро-ЦОД на крыше здания

Интернет-ЦОД

Строительство собственного ЦОД предприятия требует значительных капитальных и текущих затрат, поэтому многие крупные заказчики предпочитают вместо создания корпоративного ЦОД заключить договор SLA с коммерческими дата-центрами и получать услуги ЦОД по принципу аутсорсинга.

По данным агентства iKS-consulting, суммарная стоимость обслуживания ЦОД на 10–20 стоек превышает 3,2 млн. руб. в год. К этому необходимо добавить расходы на обучение, сертификацию сотрудников и закупку необходимых инструментов. Стоимость услуг аутсорсинга для такого дата-центра будет начинаться от 2,8 млн. руб. и расти в зависимости от уровня сервиса, заложенного в SLA.

Интернет-ЦОД предоставляет дисковое пространство для размещения веб-страниц и корпоративных сайтов и других корпоративных информационных ресурсов в сети Интернет, организации резервных IT-площадок и систем хранения корпоративной информации (размещение сервера). Интернет-ЦОД также могут быть использованы для корпоративной электронной почты. Централизованная антивирусная проверка и фильтрация нежелательной корреспонденции («спам») обеспечивает безопасность корпоративной сети и снизит затраты на интернет-трафик. Подобный аутсорсинг ИТ-систем предприятий минимизирует их финансовые и организационные затраты и позволяет сосредоточиться на основной бизнесе.

К категории Интернет-ЦОД также относятся специализированные ЦОД для хостинга Интернет-серверов различных провайдеров услуг Интернет.

ЦОД предприятий и организаций

ЦОД предприятий и организаций представляют собой частные системы Cloud Computing (Private Cloud).

Можно выделить два основных типа частных ЦОД:

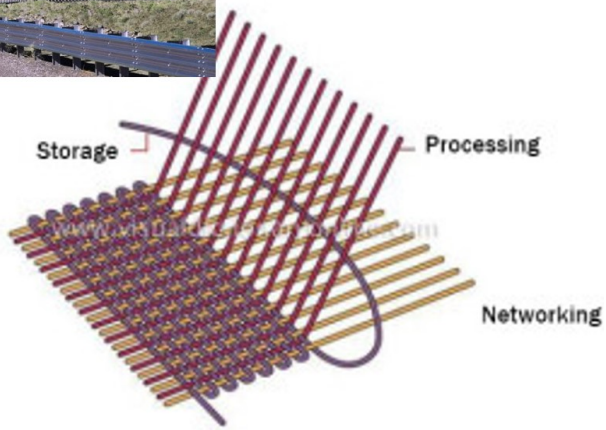
Полностью частные ЦОД, т.е. принадлежащие единственному предприятию или организации. Это могут быть коммерческие компании, правительственные учреждения или общественные организации.

ЦОД, построенные по принципу «колокации» (Co-location), которые объединяют в себе на единой инфраструктурной базе ЦОД нескольких компаний и/или организаций. Причем сам ЦОД может принадлежать «третьей стороне», т.е. компании, которая предоставляет услуги виртуализации ИТ-систем для резидентов ЦОД по принципу «внешних услуг» (Managed Service), аналогично бизнес-центрам, инфраструктурой которых (тепло- и вдоснабжение, вентиляция и кондиционирование, охрана и пр), пользуются арендаторы их площадей.

Доступ к ресурсам таких ЦОД, как правило, осуществляется по выделенным каналам передачи данных, а не через публичную сеть (Интернет).

Пример презентационного материала для проведения занятия

ЦОД(ЦхОД)



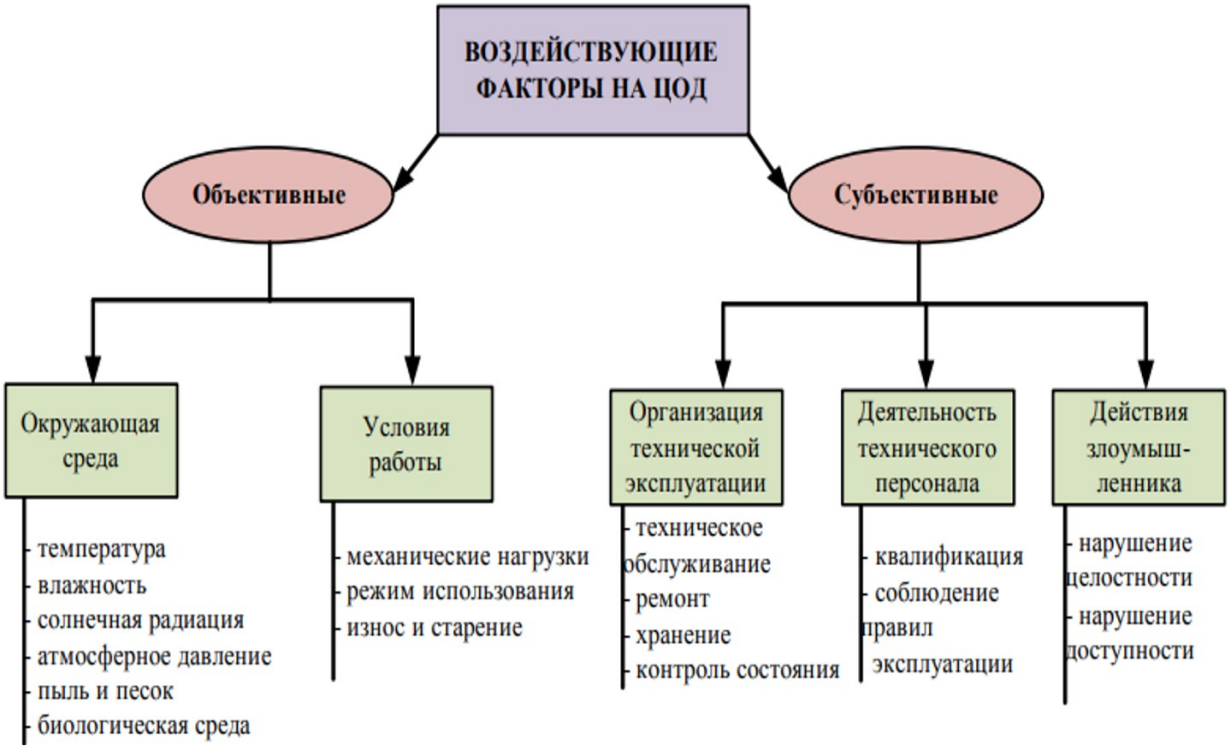
Классификация ЦОД

По назначению:

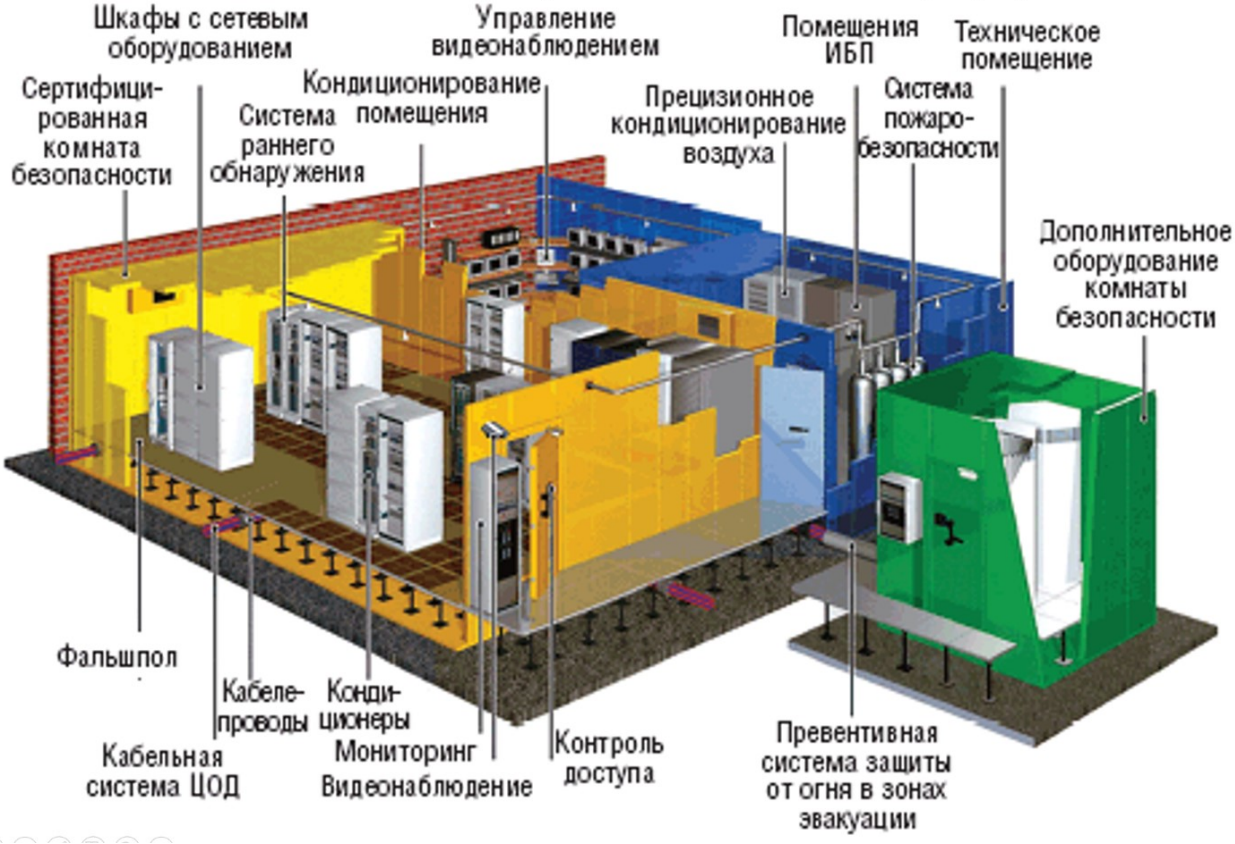
- **Основной ЦОД** – является основой информационной системы, берет на себя всю нагрузку в штатном режиме.
- **Резервный ЦОД** – заменяет основной ЦОД, в случаях выхода из строя или профилактики оборудования основного.

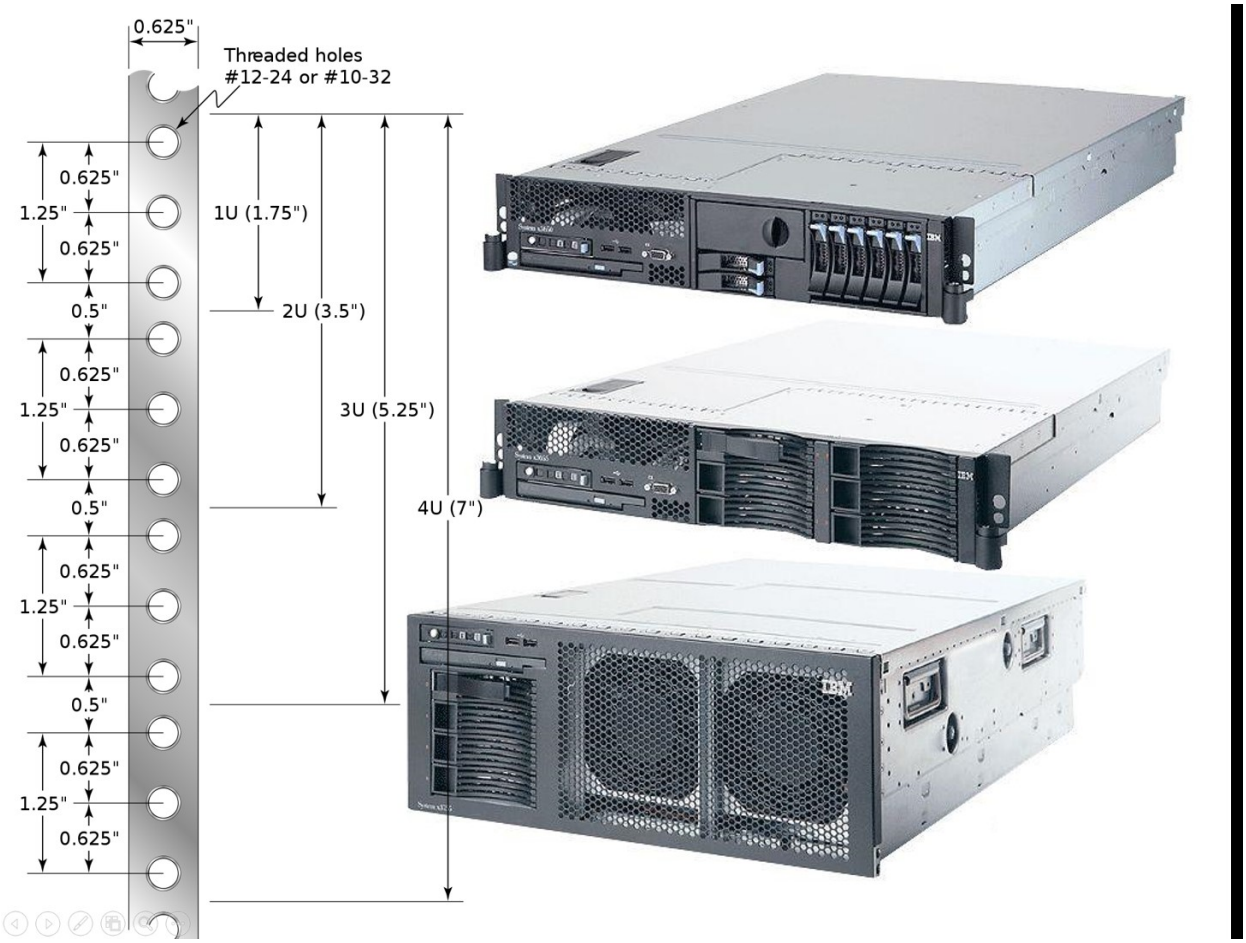
Классификация ЦОД

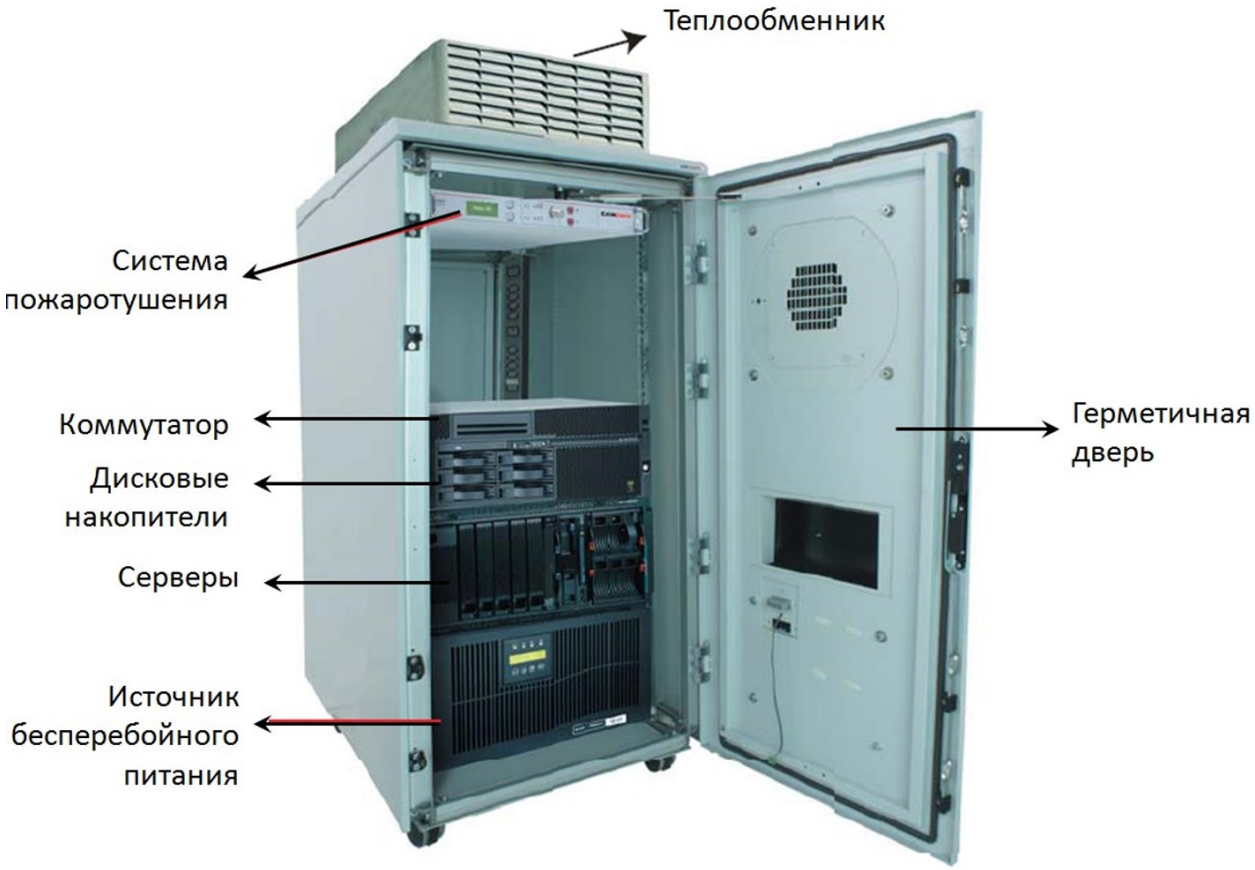
	TIER I (Уровень 1)	TIER II (Уровень 2)	TIER III (Уровень 3)	TIER IV (Уровень 4)
Год появления	1965	1970	1985	1995
Число каналов питания	Один	Один	1 активный и 1 пассивный	2 активных
Схема резервирования компонентов	N	N+1	N+1	2(N+1) или S+S
Возможность одновременной эксплуатации и ТО	НЕТ	НЕТ	ДА	ДА
Физическое разделение резервируемых компонентов	НЕТ	НЕТ	ДА	ДА
Доля фальшполов	20%	30%	80-90%	100%
Высота фальшпола	12"	13"	30-36"	30-36"
Допустимое ежегодное время простоя по вине инфраструктуры	28,8 часа	22,0 часа	1,6 часа	0,4 часа
Отношение вспомогательных площадей к площади машинного зала	10-15%	30%	80-90%	90%-100%
Бесперебойное охлаждение	Не предполагается	Не предполагается	Обязательное условие	Обязательное условие



Состав современного ЦОД




0.625"
Threaded holes
#12-24 or #10-32
0.625"
1.25"
0.625"
0.5"
0.625"
1.25"
0.625"
0.5"
0.625"
1.25"
0.625"
0.5"
0.625"
1.25"
0.625"
1U (1.75")
2U (3.5")
3U (5.25")
4U (7")

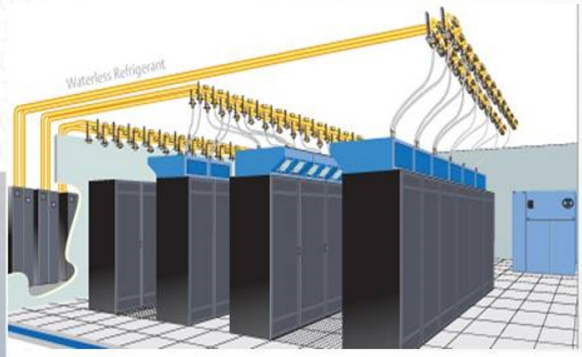

Теплообменник
Система
пожаротушения
Коммутатор
Дисковые
накопители
Серверы
Источник
бесперебойного
питания
Герметичная
дверь

Система электроснабжения.

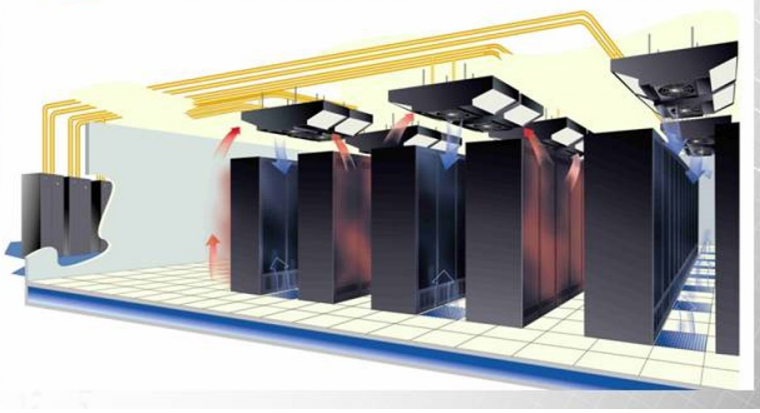
Наиболее затратный и сложный раздел проекта создания отказоустойчивого ЦОД (около 30-40% от общего бюджета на строительство)

- **система внешнего электроснабжения**
  - электроснабжение по I-ой, особой категории надежности
  - два независимых источника питания и один альтернативный
- **система внутреннего электроснабжения**
  - автоматический ввод резерва (АВР)
  - главный распределительный щит (ГРЩ)
  - Электрощитовое оборудование и Электрораспределительная сеть
- **система гарантированного электроснабжения**
  - ДГУ
- **система бесперебойного электроснабжения (СБЭ)**
  - ИБП
- **система заземления**
  - Заземление и молниезащита

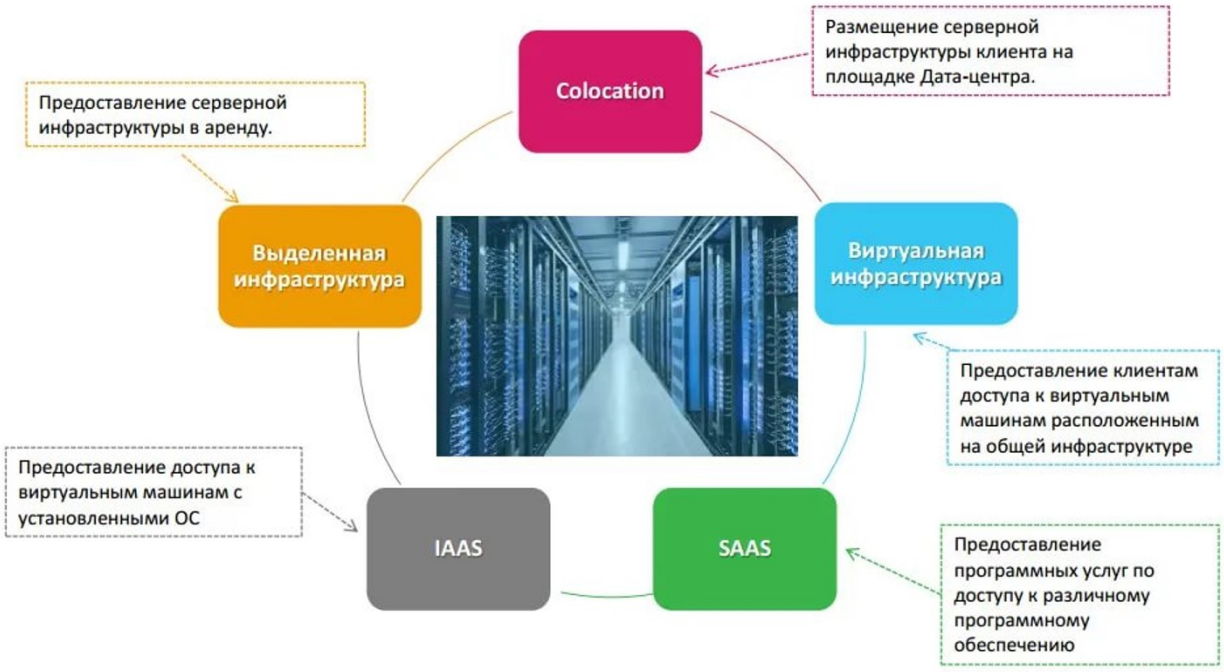
Система прецизионного кондиционирования:



- Охлаждение там где необходимо
- Работает с любыми стойками
- Хладоноситель - Фреон
- НЕТ воды в ЦОД
- Авто регулировка производительности

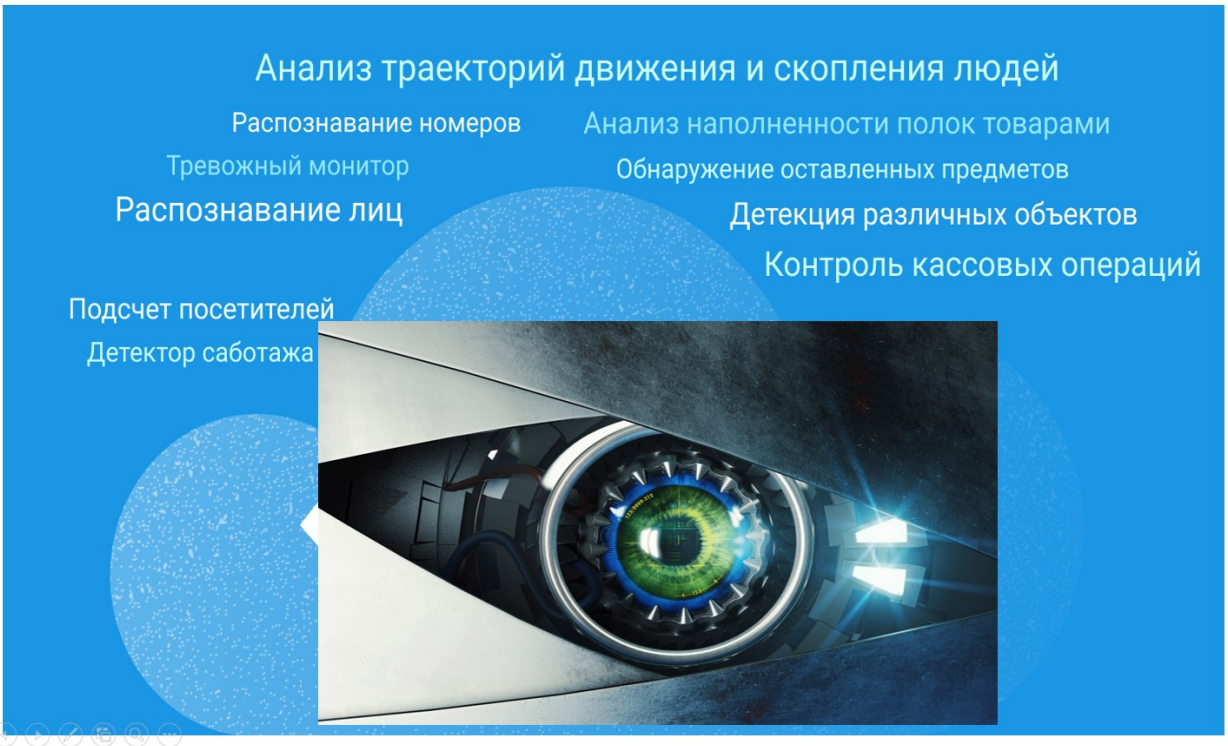


Типовые услуги ЦОД:



Vsaas (Video Surveillance as a Service)

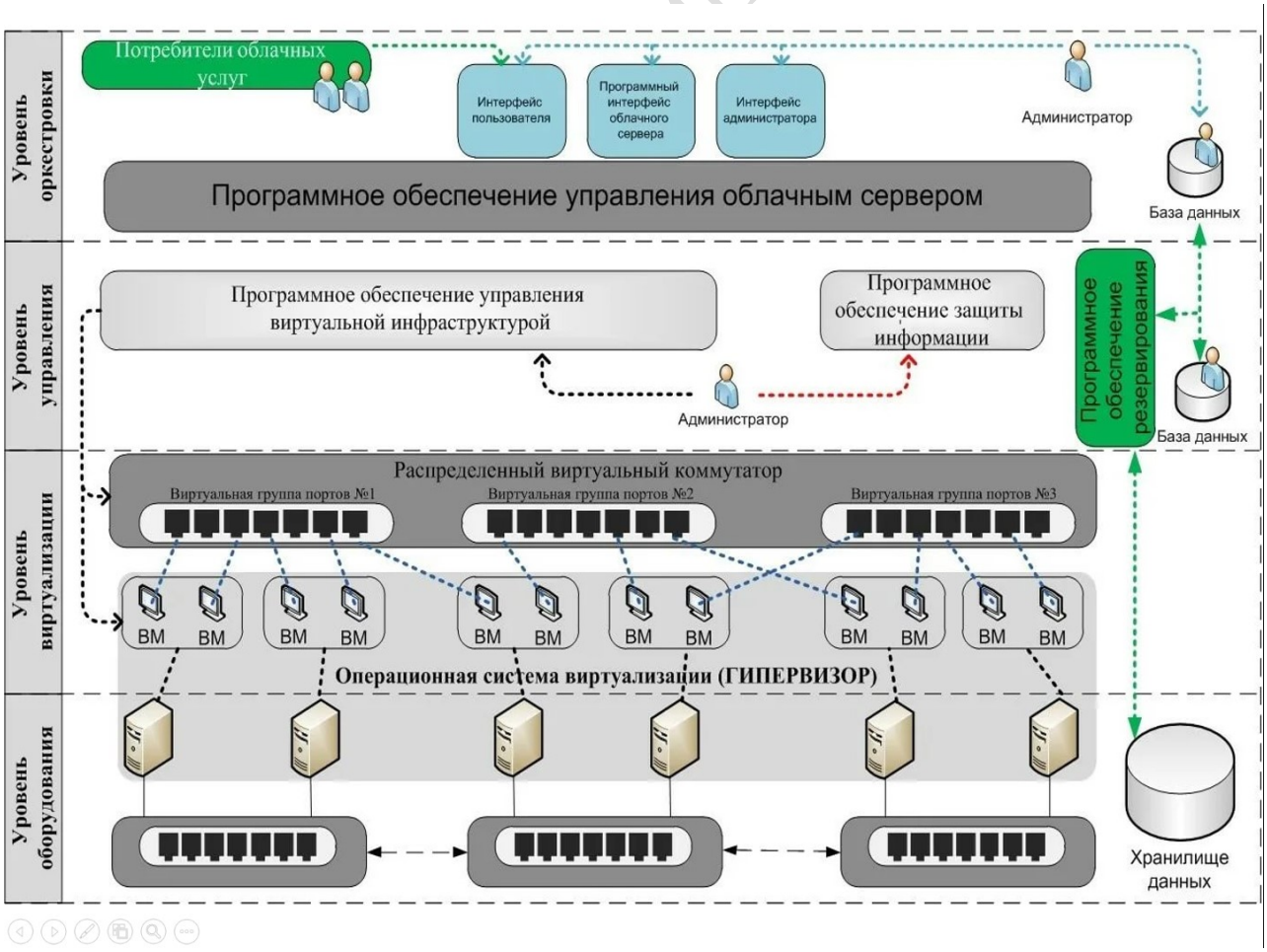
Комплекс технологий и решений, которые позволяют хранить, управлять, записывать и воспроизводить видеоматериалы с камер наблюдения из облака.



Виртуальный ЦОД

• Виртуальный ЦОД(центр обработки данных) объединяет виртуальные ресурсы в виде серверов, сетей и дисков для хранения, передачи и обработки данных. На базе ЦОД формируется IT-инфраструктура любого масштаба и сложности. По своим возможностям она не уступает аналогичным решениям с использованием физического оборудования.

• Для управления ресурсами доступен портал самообслуживания с доступом к нему с любого устройства из любой точки мира. Заказчик может самостоятельно изменять объем предоставляемых ресурсов, конфигурировать организацию виртуальных машин, создавать сетевые топологии, делать резервное копирование и так далее. Таким образом, можно эффективно контролировать все сервисы в облаке и не зависеть от технической поддержки провайдера услуг.



3 РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ УСР

Для самостоятельного изучения выделяются следующие темы:

1. Стандарты ISO серии 9000.
2. Отечественная теория и практика информационного поиска.
3. Алгоритмизация последовательности действий.
4. Изучение рынка объединенных сред разработки (Integrated Development Environment).
5. Изучение средств и возможностей операционных систем Windows.
6. Средства отображения информации.
7. Системы управления базами данных и базами знаний.
8. Современные представления о принципах действия искусственного интеллекта.
9. Глобальная телекоммуникационная система Internet.
10. Облачные сервисы предоставляемые ЦхОД.
11. Возможности средств мультимедиа и перспективы их использования.

Учебно-методическое обеспечение:

1) Рекомендуемая основная и дополнительная литература.
2) Конспект лекций по дисциплине.
3) Информация в сети Интернет.

4 ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

Тема1

1 Что такое модель OSI и для чего она нужна?
2 Общая характеристика уровней модели OSI.
3 Какие уровни содержит модель OSI.
4 Каково назначение уровней модели OSI.

Тема 2

1 Что такое каналы коммуникаций?
2 Какой сигнал называют аналоговым?
3 Какие свойства имеет аналоговый сигнал?
4 Для чего используется аналоговый сигнал?
5 Какую задачу выполняют цифровые телефонные линии?
6 Что такое модуляция? Виды модуляций.
7 Охарактеризуйте беспроводные каналы данных.
8 Для чего используется радиосвязь?
9 Виды спутниковой связи.
10 Характеристика сотовой связи.
11 Принцип работы системы TETRA.
12 Что такое каналы связи? Для чего они используются?
13 Какие каналы разграничивают системы связи (по назначению)?
14 Приведите примеры каналов связи.
15 Качественная характеристика сети ТфОП(телефонная сеть общего пользования).
16 Каковы преимущества и недостатки волоконно-оптической линии связи.

Тема 3

1 Что такое модуляция?
2 Общее описание процесса модуляции.
3 Что такое амплитудная модуляция. Формула амплитудной модуля-ции.
4 Что такое амплитудная манипуляция? Для чего она используется?
5 Что такое частотная манипуляция? Для чего она применяется?
6 Характеристика частотной модуляции.
7 Что такое линейная частотная модуляция?
8 Фазовая модуляция. Характеристика фазовой модуляции.
9 Виды фазовой манипуляции.
10 Однополосная модуляция SSB.
11 Квадратурно - амплитудная модуляция.
12 Характеристика кодово-импульсной модуляции.

Тема 4

1 Что такое модем?
2 Как происходит обмен сигналами при помощи модема?
3 Как происходит обработка переданных и принятых данных?
4 Как можно представить конструкцию модема?

5 На какие типы по исполнению делятся модемы?
6 На какие типы по принципу работы делятся модемы?
7 На какие типы по виду соединения делятся модемы?

Тема 5

1 Что такое кодирование и для чего оно используется?
2 Как кодируются дискретные сообщения?
3 Как кодируются аналоговые сообщения? Теорема Котельникова.
4 Что такое энтропия?
5 Какие существуют широко используемые двоичные коды?
6 Что необходимо для кодировки русских символов? Какая используется кодировка?
7 Характеристика синхронного кодирования.
8 Характеристика асинхронного кодирования.
9 Принцип Манчестерского кодирования.
10 Код Хемминга.
11 Что такое циклические коды? Привести пример.

Тема 6

1 Что такое сжатие данных? Для чего применяется?
2 Принцип работы алгоритма сжатия.
3 Какие бывают алгоритмы сжатия? Привести примеры.
4 Методы сжатия.
5 Предикативные методы.
6 Алгоритм MPEG. Что это? Для чего используется?
7 Описать работу метода Хаффмана.
8 Как происходит сжатие данных методом Лемпеля-Зива?
9 Каковы особенности сжатия данных методом Лемпеля-Зива?

5 ЗАДАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ

Лабораторная работа №1 Анализ информационных потоков на базе предприятия или организации.

Задание: На базе конкретного предприятия изучить структуру предприятия, провести анализ видов и типов документов данного предприятия. Используя данную информацию, дать описание структуры предприятия и его документооборота и построить схему его информационных потоков.

Лабораторная работа №2 Изучение АСУ на основе конкретного предприятия или организации.

Задание: На базе конкретного предприятия изучить структуру и функции автоматизированной системы управления предприятием, виды и назначение установленного программного обеспечения и вычислительной техники. Используя данную информацию, построить схему АСУ предприятия.

Лабораторная работа №3 Реализация различных видов программного интерфейса для взаимодействия пользователя с ПК.

Задание: Написать программу, реализующую как минимум два из интерфейсов управления программным обеспечением.

Лабораторная работа №4 Разработка простейшего редактора (текстовый, табличный, графический).

Задание: Написать программу, реализующую минимальный набор действий простейшего текстового, табличного или графического редактора с возможностью записи и чтения информации в файл.

Лабораторная работа №5 Создание презентации с использованием всех возможностей Microsoft PowerPoint.

Задание: Создать презентацию, включающую в себя ряд слайдов разного типа (включая титульный, текст и графика, вставку аудио- и видеоряда), применив к объектам форматирование, эффекты анимации и шаблоны оформления слайдов.

Лабораторная работа №6 Изучение возможностей Microsoft Access на примере создания простой базы данных.

Задание: Создать простую базу данных, приведенную к 3-й нормальной форме, включающую в себя несколько таблиц данных, форм ввода (редактирования) данных, создать запросы к базе и несколько отчетов.

Лабораторная работа №7 Создание макросов в Microsoft Excel. Импорт данных в Excel при помощи макросов.

Задание: Создать простую книгу Excel и записать в нее макрос по выполнению произвольных действий над данными в Excel. Осуществить просмотр кода макроса и его выполнение. Подготовить вариант текстового файла для импорта в Excel. Осуществить импорт файла в книгу Excel. Разработать макрос по импорту файлов в Excel (обязательное наличие возможности выбора файлов или их имен при импорте).

Отчет должен содержать пример работы записанных макросов, часть импортируемого в Excel файла, листинг первого (произвольного) макроса, листинг макроса импорта.

Лабораторная работа №8.

Форматирование данных в Excel и построение диаграмм при помощи макросов

Задание: Разработать набор макросов по форматированию текста в Excel и построению на основании отформатированных данных ряда диаграмм различных типов (обязательные виды диаграмм – гистограмма и круговая, плюс два произвольных вида).

Отчет должен содержать перечень действий по форматированию с примерами их реализации, реализуемых макросами, вид данных до форматирования и после форматирования, примеры построенных диаграмм, листинг макросов.

Лабораторная работа №9.

Написание криптографической защиты информации. Шифрование данных

Задание: Выбрать алгоритм шифрования данных и написать программу, которая бы его реализовывала. На вход программе подается некоторый (произвольный) файл, который шифруется согласно выбранному алгоритму. На выходе программы формируется зашифрованный файл.

Отчет должен содержать описание алгоритма шифрования (в виде блок схемы-для простых алгоритмов), набор исходных и выходных файлов, примеры работы программы в виде скриншотов, листинг программы. В выводе к работе оценить алгоритм шифрования.

Лабораторная работа №10. Написание криптографической защиты информации. Дешифрование данных

Задание: На основании алгоритма шифрования данных написать программу, которая бы обеспечивала дешифрацию закодированных данных. На вход программе подается произвольный зашифрованный файл, который дешифруется согласно выбранному алгоритму. На выходе программы формируется дешифрованный файл.

Отчет должен содержать описание алгоритма дешифрования (в виде блок-схемы для простых алгоритмов), набор исходных и выходных файлов,

примеры работы программы в виде скриншотов, листинг программы. В выводе к работе оценить алгоритм дешифрования.

Лабораторная работа №11. Разработка защиты программного обеспечения с использованием программных средств

Задание: Изучить все популярные системы защиты ПО при помощи программных средств. Разработать и реализовать алгоритм защиты произвольной программы на уровне программных средств, т.е. алгоритм защиты должен использовать либо поддержку серийный номеров или ключей авторизации (для полнофункциональных программ), либо привязку к программным счетчикам (для программ с ограниченным сроком действия или набором действий), либо многоступенчатую регистрацию или любой другой способ.

Отчет должен содержать описание алгоритма или принципа защиты программы (в виде блок-схемы для простых алгоритмов), примеры работы программы в виде скриншотов, листинг программы. В выводе к работе оценить алгоритм защиты.

Лабораторная работа №12. Разработка защиты программного обеспечения при помощи аппаратных средств

Задание: Разработать и реализовать алгоритм защиты произвольной программы на уровне аппаратных средств, т.е. алгоритм защиты должен использовать произвольный элемент аппаратной базы (ID номер жесткого диска, BIOS различных плат, аппаратные hasp- или flash-ключи и многое другое).

Отчет должен содержать описание алгоритма или принципа защиты программы (в виде блок-схемы для простых алгоритмов), примеры работы программы в виде скриншотов, листинг программы. В выводе к работе оценить алгоритм защиты.

Лабораторная работа №13. Анализ программного обеспечения для работы и защиты информации в Internet

Задание: Необходимо изучить возможности программного пакета, который осуществляет работу в Internet либо обеспечивает защиту информации в Internet.

Отчет должен содержать описание возможностей изучаемого программного продукта и параметры его настройки, исследование его уязвимостей. В выводе к работе оценить изученный программный продукт с точки зрения его безопасности и надежности при работе в сети Internet.

6 ТЕСТОВЫЕ ЗАДАНИЯ (примеры)

1. Скорость передачи информации - это …
   - среднее количество информации, получаемое на выходе канала в единицу времени.
   - среднее количество символов, передаваемое по каналу за единицу времени.
   - максимальная скорость передачи информации по каналу связи в единицу времени.
2. Расставьте этапы обработки информации в правильном порядке
   Очистка, первичная обработка и приведение к унифицированному виду
   - Второй этап
     Систематизация и организация хранения накопленных данных, для последующего использования
   - Третий этап
     Формирование отчета по конкретной тематике
   - Пятый этап
     Первоначальный сбор из внешних источников
   - Первый этап
     Глубокий анализ информации, систематизация и получение знаний.
   - Четвертый этап
3. Дополните определение: _____- это предметно-ориентированное, привязанное ко времени и неизменяемое собрание данных для поддержки процесса принятия управляющих решений.
4. Сохранность - это состояние документа, программы или технических средств, характеризуемое
   - возможностью использования в любое время
   - возможностью открыть их
   - степенью удержания их эксплуатационных свойств
   - скоростью доступа к ним
5. Допишите определение: Информационная система – совокупность технического, программного и информационного обеспечения, а также персонала, предназначенная для того, чтобы своевременно обеспечивать людей надлежащей информацией
6. Отметьте возможные операции с данными в хранилище:
   - извлечение
   - преобразование
   - загрузка
   - анализ
   - представление результатов анализа
   - изменение
   - очистка
   - шифрование

7. Определите вид OLAP системы по описанию: этот вид OLAP системы работает с реляционными базами данных. Обращение к данным осуществляется напрямую в реляционную базу данных. Данные хранятся в виде реляционных таблиц
   - HOLAP
   - ROLAP
   - MOLAP
   - DOLAP
8. Допишите определение: Жизненный цикл информационной системы является производной жизненного цикла _____ (1 слово)
9. Выберите существующие модели жизненного цикла информационных систем
   - спиральная
   - круговая
   - ячеистая
   - каскадная
   - поэтапная
   - пошаговая
10. «Набор инструментов и методов программной инженерии для проектирования ПО, позволяющий обеспечить высокое качество программ, отсутствие ошибок и простоту обслуживания программных продуктов. При проектировании этими средствами ПО строится из отдельных блоков». Речь идет о:
   - Метолодогии CMM
   - CASE-средствах
   - Методологии CMMI
   - RASE стредствах
11. Система класса _____ (аббревиатура)- это корпоративная информационная система для автоматизации планирования, учета, контроля и анализа всех основных бизнес-процессов и решения бизнес задач в масштабе предприятия (организации).
12. практический инструмент, созданный в рамках процессного подхода к описанию деятельности проектной организации, в частности, организации, разрабатывающей информационные системы, демонстрирует методология _____ Допишите.
13. Допишите определение: _____ (Аббревиатура) – специализированно здание для размещения серверного и сетевого оборудования и подключения абонентов к каналам сети интернет.
14. На какие две части разделяется любой центр обработки данных
   - информационную и техническую
   - коммерческого и некоммерческого использования.
   - цифровую и аналоговую
   - внутреннюю и внешнюю
15. Какие четыре действия являются основными для СУБД

- Поиск, вставка, удаление, замена элемента
- Копирование, вставка, удаление, замена
- Копирование, вырезание, вставка, удаление

16. Технология непрерывной информационной поддержки жизненного цикла продукта называется - _____ (англ. аббревиатура)
17. Допишите определение: _____ - способ сокрытия исходного смысла сообщения или другого документа, обеспечивающей искажение его первоначального содержимого и невозможности прочтения без скециального ключа.
18. Допишите определение. _____ — какая-либо система преобразования текста с секретом (ключом) для обеспечения секретности передаваемой информации.
19. Выберите правильные виды аналоговой модуляции сигналов:
    - АМ
    - ВМ
    - ЧМ
    - ЛЧМ
    - ФМ
    - КМ
    - СКМ
20. Выберите правильные виды цифровой модуляции сигналов:
    - АМн
    - ПАМ
    - ФМн
    - КАМ
    - ФАМ
    - ЧМн
    - Треллис-модуляция
21. Выберите правильные виды импульсной модуляции сигналов:
    - АИМ
    - ДМ
    - АМ
    - ИКМ
    - ШКМ
    - ШИМ
    - ЧИМ
    - ФИМ

6 Учреждение образования
«Гомельский государственный университет имени Франциска Скорины»

УТВЕРЖДАЮ
Проректор по учебной работе
УО «ГГУ им. Ф. Скорины»

_____ И.В. Семченко
(подпись)

_____
(дата утверждения)

Регистрационный № УД-_____/уч.

ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Учебная программа государственного компонента по учебной дисциплине для специальности:

1 – 53 01 02 **Автоматизированные системы обработки информации.**
(код специальности) **(наименование специальности)**

2021 г.

Учебная программа составлена на основе: образовательного стандарта ОСВО 1-53 01 02-2021 г. и типового учебного плана, регистрационный № I 53-1-010/пр-тип., дата утверждения 08.04.2021

СОСТАВИТЕЛЬ:

В.Н. Кулинченко, ст. преподаватель кафедры АСОИ

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой автоматизированных систем обработки информации (протокол № *9* от 13.04.2021);

Научно-методическим советом Учреждения образования «ГГУ имени Ф.Скорины» (протокол № 6 от 05.05.2021).

Пояснительная записка

Изучение дисциплины государственного компонента «Информационные системы и технологии» предусмотрено типовым учебным планом подготовки специалистов специальности I-53 01 02 – «Автоматизированные системы обработки информации» в соответствии с требованиями образовательного стандарта высшего образования первой ступени.

Актуальность изучения дисциплины обусловлена тем, что информационные системы и технологии стали характеризовать уровень развития общества, а также возможность каждого будущего специалиста высшей категории использовать современные информационные технологии в качестве механизма для решения сложных задач в своей профессиональной деятельности.

Целью изучения курса является подготовка студентов по основам построения и эксплуатации информационных систем и технологий в области управления и обработки информации.

Основные задачи изучения дисциплины:

– формирование представления об информационной технологии как специфической системе, требующей комплексного подхода, использования средств системного анализа;

– приобретение знаний о возможностях, методах, моделях и средствах информационных технологий;

– приобретение навыков практической работы со средствами современных информационных технологий.

Для изучения курса «Информационные системы и технологии» необходимы базовые знания по следующим дисциплинам: «Основы алгоритмизации и программирования». В свою очередь учебная дисциплина «Информационные системы и технологии» является базой для таких учебных дисциплин как «Базы и банки данных» (компонент учреждения высшего образования).

В результате изучения учебной дисциплины «Информационные системы и технологии» формируются следующие компетенции:

базовые профессиональные:

получать, обрабатывать и анализировать информацию, обеспечивать ее хранение.

В результате изучения учебной дисциплины обучающийся должен:

знать:

– основные понятия информационных систем и технологий;

– методологические основы информационных систем и технологий;

– концептуальные основы информационных технологий;

– основные методы описания информационных процессов;

– технологии и средства поиска информации;

– методы преобразования и оценки качества преобразования информации;

– основные криптографические алгоритмы;

– способы восстановления данных;

– специфику технологий обработки и хранения данных;

– иметь четкое представление о развитии современных информационных технологий в своей предметной/профессиональной деятельности;

– знать о методах и средствах, которые основаны на базе современных информационных технологий, для решения задач в своей предметной/профессиональной области;

– иметь четкое представление о программном обеспечении современных информационных технологий.

уметь:

– применять информационные методы для описания объектов автоматизированных информационных систем;

– работать с современными системами поиска информации;

– применять средства мультимедиа;

– работать с пакетами для сжатия и шифрования данных;

– передавать данные через сеть

– освоить основные приемы данной дисциплины (на персональных компьютерах), также ознакомиться с техническим обеспечением современных информационных систем и технологий;

– ознакомиться с проблемами защиты информации непосредственно на персональных компьютерах, в компьютерных сетях.

Дисциплина государственного компонента «Информационные системы и технологии» изучается студентами 1 курса дневной формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации»; студентами 1 курса заочной формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации»; студентами 1 курса заочной интегрированной со средним специальным образованием формы обучения специальности I-53 01 02 – «Автоматизированные системы обработки информации».

Дневная форма обучения: всего часов по плану-108, аудиторное количество часов – 108; из них: лекционных занятий – 32 (в том числе УСР 6), лабораторных работ – 24.

Форма отчётности – экзамен в 1 семестре.

Заочная форма обучения: всего часов по плану-108, аудиторное количество часов – 14, из них: лекционных занятий –8, лабораторных работ – 6.

Форма отчётности – экзамен во 2 семестре.

Заочная форма обучения (интегрированная на основе среднего специального образования): всего часов по плану-108, аудиторное

количество часов – 14, из них: лекционных занятий – 8, лабораторных работ – 6.

Форма отчётности – экзамен во 2 семестре.

Содержание учебного материала

ВВЕДЕНИЕ

Вещественные, энергетические и информационные процессы в современном обществе. Понятие «технология». Основные аспекты понятия технология. Информационная технология. Особенности понятия информационной технологии. Современные информационные технологии и их особенности. Информационная технология и информационная система. Основные виды информационных систем.

Задачи, содержание и место курса в инженерной подготовке. Связь курса с другими дисциплинами специальности.

РАЗДЕЛ 1. МЕТОДОЛОГИЧЕСКИЙ БАЗИС ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Взаимосвязь информационной технологии как научной дисциплины с другими научными направлениями. Понятие системы. Основные свойства системы. Структура, архитектура и цель системы. Основные аспекты структуры сложной системы. Разработка архитектуры информационной системы.

Качество системы. Стандарты ISO серии 9000. Назначение и особенности стандартов серии 9000. Модель «уровней зрелости» CMM. Основные модели CMM. Стандарт CMMI. Основные характеристики уровней CMM.

РАЗДЕЛ 2. КОНЦЕПТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Концепция открытых систем. Методологические основы открытых систем. Основополагающие документы, определяющие концепцию открытых систем.

Эталонная модель OSI. Основные понятия модели OSI. Уровни OSI. Основные задачи и выполняемые функции. Понятие стека протоколов.

Общие сведения о стандартах в области информационных технологий. Роль стандартов в области информационных технологий. Уровни и виды стандартов.

Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека. Методология современных информационных систем. Информационные технологии корпоративных и государственных учреждений. Системы класса ERP. Корпоративные порталы. CALS-технологии.

РАЗДЕЛ 3. ОСНОВНЫЕ ПОДХОДЫ И МЕТОДЫ ОПИСАНИЯ ИНФОРМАЦИОННЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ

Информация и данные. Виды и свойства информации. Сигналы и знаки. Классификация сигналов. Математические модели сигналов. Теория сигналов, семиотика, теория информации. Основные направления семиотики.

Синтаксические, семантические и прагматические направления и меры информации. Роль классической теории информации в становлении ряда прикладных дисциплин и развитии автоматизированных информационных технологий.

РАЗДЕЛ 4. ОСНОВЫ КЛАССИЧЕСКОЙ ТЕОРИИ ИНФОРМАЦИИ

Количество информации при конечном числе равновозможных исходов. Мера Хартли. Количество информации как случайная величина. Энтропия. Основные свойства энтропии. Среднее количество взаимной информации (дискретный случай). Энтропия объектов с непрерывным множеством состояний. Среднее количество взаимной информации (непрерывный случай).

Информационные характеристики источников сообщений. Источники дискретных сообщений. Энтропия источника дискретных сообщений. Понятие избыточности источника сообщений. Скорость создания информации источником дискретных сообщений. Источники непрерывных сообщений. Информационные характеристики источников непрерывных сообщений.

Информационные характеристики каналов связи. Понятие канала связи. Понятие скорости передачи и пропускной способности канала.

РАЗДЕЛ 5. ОСНОВЫ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Тема 5.1. ВОСПРИЯТИЕ ИНФОРМАЦИИ

Процесс восприятия информации и его особенности. Основные этапы восприятия информации. Первичное восприятие, обнаружение, распознавание, анализ информации. Схема процесса восприятия

информации. Физический, морфологический, синтаксический и семантический аспекты восприятия.

Технология поиска информации. Поиск информации в Интернет. Поисковые машины, метапоисковые средства, онлайновые энциклопедии и справочники. Современные поисковые порталы. Извлечение ключевых слов из текстовых материалов в MS Word. Программы-экстракторы. Системы автоматического анализа текста.

Тема 5.2. ПРЕОБРАЗОВАНИЕ ИНФОРМАЦИИ

Цели и виды преобразования информации. Редукция, кодирование, модуляция. Дискретизация сигнала во времени. Основные методы дискретизации сигнала. Оценка погрешности дискретизации. Квантование сигнала по уровню. Дисперсия шума квантования. Цифровое представление информации. Двоичная, восьмеричная и шестнадцатеричная формы представления.

Кодирование информации. Статистическое и помехоустойчивое кодирование. Шифрование данных. Основные криптографические методы. Симметричные алгоритмы шифрования. Алгоритмы шифрования с открытым ключом. Алгоритм RSA. Алгоритм Эль-Гамаля.

Модуляция. Амплитудная, частотная и фазовая модуляции. Понятие спектра сигнала. Спектр колебаний модулированного сигнала. Импульсная модуляция.

Тема 5.3. ПЕРЕДАЧА ИНФОРМАЦИИ

Передача информации и коммуникационные технологии. Коммуникационные сети. Принципы построения цифровых каналов. Информационная модель канала связи.

Пропускная способность дискретного канала без шума. Основная теорема Шеннона для дискретного канала без шума. Эффективное кодирование. Коды Шеннона-Фано и Хаффмена. Современные методы сжатия данных. Пропускная способность дискретного канала с шумом. Основная теорема Шеннона для дискретного канала с шумом. Пропускная способность непрерывного канала с шумом.

Методы повышения помехоустойчивости передачи данных. Помехи. Модели ошибок в реальных каналах. Основные методы повышения помехоустойчивости передачи данных. Методы оптимального приема сигналов. Бинарное обнаружение. Критерии оптимальности бинарного обнаружения. Структура оптимального приемника.

Помехоустойчивое кодирование. Принципы помехоустойчивого кодирования. Принципы построения корректирующих кодов. Понятие группы и поля. Групповые коды. Коды Хэмминга. Порождающая и проверочная матрицы групповых кодов. Циклические коды. Принципы

повышения помехоустойчивости передачи данных, основанные на использования обратной связи между выходом и входом канала.

Тема 5.4. ОБРАБОТКА ИНФОРМАЦИИ

Понятие обработки информации. Обработка данных и переработка информации. Основные виды обработки информации. Технологический процесс обработки данных. Технологическая сеть обработки данных. Типовые операции обработки данных. Сбор, регистрация, сортировка, поиск и выдача информации.

Тема 5.5. ХРАНЕНИЕ ИНФОРМАЦИИ

Физические основы хранения информации. Типы физических носителей. Специфика хранения информации. Интерфейсы физических носителей. Системы кодирования информации на физических носителях.

Тема 5.6. ТЕХНОЛОГИИ ХРАНЕНИЯ ИНФОРМАЦИИ

Основные аспекты хранения информации. Базы данных. Хранилища данных и их классификация. Технология OLAP. Основные сферы применения технологии OLAP-кубов данных. Хранилища данных. ЦхОДы и их классификация.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ (Дневная форма обучения)

Номер раздела, темы, занятия	Название раздела, темы, занятия; перечень изучаемых вопросов	Количество аудиторных часов				Иное	Формы контроля знаний
		лекции	практические занятия	лабораторные занятия	Количество часов УСР		
1	2	3	4	5	6	7	9
	ВВЕДЕНИЕ. (2 Ч.)	2					
1	*РАЗДЕЛ 1 МЕТОДОЛОГИЧЕСКИЙ БАЗИС ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ (10 Ч.)*	**4**		**4**	**2**	Методическое пособие	
1.1	*Информационная технология как сложная система (6 ч.)* 1 Понятие системы. 2 Основные свойства системы. 3 Структура, архитектура и цель системы. 4 Разработка архитектуры информационной системы	2		4		[1] [3] [7]	
1.2	**Качество системы (2 ч.)** **1** Роль стандартов в области информационных технологий. 2 Уровни и виды стандартов. 3 Стандарты ISO серии 9000. Назначение и особенности стандартов серии 9000.				2	[2] [5]	
1.3	**Модель «уровней зрелости» CMM (2 ч.)** 1 Стадии жизненного цикла проекта. 2 Основные модели CMM. 3 Стандарт CMMI. 4 Основные характеристики уровней CMM.	2					
2	*РАЗДЕЛ 2. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ (8 Ч.)*	**4**		**4**		Методическое пособие	

2.1	**Концепция открытых систем. (4 ч.)** 1 Методологические основы открытых систем. 2 Основополагающие документы, определяющие концепцию открытых систем. 3 Основные понятия модели OSI. 4 Уровни OSI. Основные задачи и выполняемые функции. 5 Понятие стека протоколов.	2		2		[2] [3] [4]	Защита лабораторных работ
2.2	**Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека (4 ч.)** 1 Методология современных информационных систем. 2 Информационные технологии корпоративных и государственных учреждений. 3 Системы класса ERP. 4 Корпоративные порталы. CALS-технологии.	2		2		[1] [4] [8]	Защита лабораторных работ
3	Раздел 3. Основные подходы и методы описания информационных явлений и процессов (4 ч.)	**2**			**2**	Методическое пособие	
3.1	**Информация и данные (2 ч.)** 1 Виды и свойства информации. 2 Сигналы и знаки. 3 Классификация сигналов. 4 Математические модели сигналов. 5 Теория сигналов, семиотика, теория информации. 6 Основные направления семиотики.	2				[1] [3] [7]	
3.2	**Основные меры информации (2 ч.)** 1 Синтаксические, семантические и прагматические направления и меры информации. 2 Роль классической теории информации в становлении ряда прикладных дисциплин и развитии автоматизированных информационных технологий.				2	[2] [3] [10]	
4	Раздел 4. Основы классической теории информации (4 ч.)	**4**				Методическое пособие	
4.1	**Количество информации при конечном числе равновозможных исходов. Мера Хартли (4 ч.)** 1 Количество информации как случайная величина. 2 Энтропия. 3 Основные свойства энтропии. 4 Среднее количество взаимной информации (дискретный случай).	2				[3] [7] [12]	Защита лабораторных работ

	5 Энтропия объектов с непрерывным множеством состояний. 6 Среднее количество взаимной информации.						
4.2	**Информационные характеристики источников и каналов связи (2 ч.)** 1 Источники дискретных сообщений. 2 Понятие избыточности источника сообщений. 3 Информационные характеристики источников непрерывных сообщений. 4 Понятие канала связи. **5** Понятие скорости передачи и пропускной способности канала.	2					
5	Раздел 5 Основы информационных процессов (26 ч.)	**8**		**16**	**2**	Методическое пособие	
5.1	Восприятие информации (4 ч.) 1 Процесс восприятия информации и его особенности. 2 Основные этапы восприятия информации. 3 Первичное восприятие, обнаружение, распознавание, анализ информации. 4 Схема процесса восприятия информации. 5 Физический, морфологический, синтаксический и семантический аспекты восприятия машинных интерфейсов.			2	2	[2] [4] [8]	Реферативная работа
5.2	**Преобразование информации (6 ч.)** 1 Цели и виды преобразования информации. 2 Редукция, кодирование, модуляция. Дискретизация сигнала во времени. 3 Кодирование информации. 4 Шифрование данных. 5 Основные криптографические методы.	2		4		[1] [2] [11]	Защита лабораторных работ Реферативная работа
5.3	**Передача информации (6 ч.)** 1 Передача информации и коммуникационные технологии. 2 Коммуникационные сети. Принципы построения цифровых каналов. 3 Пропускная способность дискретного канала без шума. 4 Основная теорема Шеннона для дискретного канала без шума. 5 Современные методы сжатия данных 6 Методы повышения помехоустойчивости передачи данных. 7 Принципы построения корректирующих кодов. 8 Помехоустойчивое кодирование. Коды Хэмминга.	2		4		[1] [4] [8]	Защита лабораторных работ Реферативная работа
5.4	Обработка информации (6 ч.) 1 Понятие обработки информации. 2 Обработка данных и переработка информации. 3 Основные виды обработки информации. 4 Типовые операции обработки данных. 5 Сбор, регистрация, сортировка, поиск и выдача информации.	2		4		[2] [10]	Защита лабораторных работ

5.5	Хранение информации (4 ч.) 1 Физические основы хранения информации. 2 Типы физических носителей. 3 Интерфейсы физических носителей. 4 Специфика хранения информации.	2		2		[2] [7] [8]	Защита лабораторных работ
5.6	Технологии хранения информации (2 ч.) 1 Основные аспекты хранения информации. 2 Хранилища данных. ЦхОДы и их классификация. 3 Планарные модели хранения информации. 4 OLAP-кубы данных.	2				[2] [7] [8]	Защита лабораторных работ
	Всего	**26**		**24**	**6**		

Старший преподаватель В.Н. Кулинченко

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ (заочная форма обучения, заочная интегрированная на основе среднего специального образования форма обучения, дистанционная форма обучения)

Номер раздела, темы, занятия	Название раздела, темы, занятия; перечень изучаемых вопросов	Количество аудиторных часов				Иное	Формы контроля знаний
		лекции	практические занятия	лабораторные занятия	Количество часов УСР		
1	2	3	4	5	6	7	9
2	*РАЗДЕЛ 2. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ*	**2**		**2**		Методическое пособие	
2.1	**Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека** 1 Информационные технологии корпоративных и государственных учреждений. 2 Системы класса ERP, BI. 3 Корпоративные порталы. CALS-технологии.	2		2		[1] [4] [8]	Защита лабораторных работ
3	Раздел 3. Основные подходы и методы описания информационных явлений и процессов	**2**				Методическое пособие	
3.1	**Информация и данные (2 ч.)** 1 Виды и свойства информации. 2 Сигналы и знаки. 3 Классификация сигналов. 4 Математические модели сигналов. 5 Теория сигналов, семиотика, теория информации. 6 Понятие канала связи. 7 Понятие скорости передачи и пропускной способности канала.	2				[1] [3] [7]	
5	Раздел 5 Основы информационных процессов	**4**		**4**		Методическое	

						пособие	
5.1	**Передача и преобразование информации** 1 Цели и виды преобразования информации. 2 Редукция, кодирование, модуляция. Дискретизация сигнала во времени. 3 Кодирование информации. 4 Шифрование данных. 5 Основные криптографические методы. 6 Передача информации и коммуникационные технологии. 7 Коммуникационные сети. Принципы построения цифровых каналов. 8 Современные методы сжатия данных 9 Методы повышения помехоустойчивости передачи данных.	2		2		[1] [2] [11]	Защита лабораторных работ
5.2	Обработка и хранение информации 1 Понятие обработки информации. 2 Обработка данных и переработка информации. 3 Основные виды обработки информации. 4 Типовые операции обработки данных. 5 Сбор, регистрация, сортировка, поиск и выдача информации. 6 Физические основы хранения информации. 7 Типы физических носителей. 8 Специфика хранения информации. 9 Основные аспекты хранения информации. 10 Технологии хранения данных. ЦхОДы и их классификация.	2		2		[2] [10]	Защита контрольной работы
	Всего	**8**		**6**			**экзамен**

Старший преподаватель В.Н. Кулинченко

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА (Дистанционная форма обучения)

Номер раздела, темы, занятия	Название раздела, темы, занятия; перечень изучаемых вопросов	Количество аудиторных часов				Иное	Формы контроля знаний
		лекции	практические занятия	лабораторные занятия	Количество часов УСР		
1	2	3	4	5	6	7	9
2	*РАЗДЕЛ 2. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ*	**2**				Методическое пособие	
2.1	**Классификация информационных технологий по укрупненным видам и сферам информационной деятельности человека** 1 Информационные технологии корпоративных и государственных учреждений. 2 Системы класса ERP. 3 Корпоративные порталы. CALS-технологии.	2		2		[1] [4] [8]	Защита лабораторных работ
3	Раздел 3. Основные подходы и методы описания информационных явлений и процессов	**2**				Методическое пособие	
3.1	**Информация и данные (2 ч.)** 1 Виды и свойства информации. 2 Сигналы и знаки. 3 Классификация сигналов. 4 Математические модели сигналов. 5 Теория сигналов, семиотика, теория информации. 6 Понятие канала связи. 7 Понятие скорости передачи и пропускной способности канала.	2				[1] [3] [7]	
5	Раздел 5 Основы информационных процессов	**4**		**4**		Методическое пособие	

5.1	**Передача и преобразование информации** 1 Цели и виды преобразования информации. 2 Редукция, кодирование, модуляция. Дискретизация сигнала во времени. 3 Кодирование информации. 4 Шифрование данных. 5 Основные криптографические методы. 6 Передача информации и коммуникационные технологии. 7 Коммуникационные сети. Принципы построения цифровых каналов. 8 Современные методы сжатия данных 9 Методы повышения помехоустойчивости передачи данных.	2		2		[1] [2] [11]	Защита лабораторных работ
5.2	Обработка и хранение информации 1 Понятие обработки информации. 2 Обработка данных и переработка информации. 3 Основные виды обработки информации. 4 Типовые операции обработки данных. 5 Сбор, регистрация, сортировка, поиск и выдача информации. 6 Физические основы хранения информации. 7 Типы физических носителей. 8 Специфика хранения информации. 9 Основные аспекты хранения информации. 10 Технологии хранения данных. ЦхОДы и их классификация.	2		2		[2] [10]	Защита контрольной работы
	Всего	**8**		**6**			**экзамен**

Старший преподаватель В.Н. Кулинченко

ИНФОРМАЦИОННО - МЕТОДИЧЕСКАЯ ЧАСТЬ

ПЕРЕЧЕНЬ ТЕМ ЛАБОРАТОРНЫХ РАБОТ

1. Анализ информационных потоков на базе предприятия или организации
2. Реализация различных видов программного интерфейса для взаимодействия пользователя с ПК
3. Разработка простейшего редактора (текстовый, табличный, графический)
4. Создание презентации с использованием всех возможностей Microsoft PowerPoint
5. Изучение возможностей Microsoft Access на примере создания простой базы данных
6. Создание макросов в Microsoft Excel. Импорт данных в Excel при помощи макросов
7. Форматирование данных в Excel и построение диаграмм при помощи макросов
8. Написание криптографической защиты информации. Шифрование данных
9. Написание криптографической защиты информации. Дешифрование данных
10. Разработка защиты программного обеспечения с использованием программных средств
11. Разработка защиты программного обеспечения при помощи аппаратных средств
12. Анализ программного обеспечения для комплексной защиты информации в Internet

ПЕРЕЧЕНЬ НЕОБХОДИМОГО ОБОРУДОВАНИЯ И КОМПЬЮТЕРНЫХ ПРОГРАММ

1. Класс современных персональных ЭВМ.
2. Программные средства лабораторного практикума курса «Информационные системы и технологии»

8 РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная

1. Каймин, В. А. Информатика: учебник / В.А. Каймин. - М.: Инфра-М, 2010.—272 с.

2.Вудкок Дж. Современные информационные технологии совместной работы: пер. с англ. / Дж.Вудкок. – М.: Русская редакция, 2007. – 256 с.

3. Крейнак, Дж. Интернет. Серия Энциклопедия: пер. с англ. / Дж.Крейнак. – СПб.: Питер Паблишинг, 2011. -- 555 с.

5. Федоров, А. Г. Базы данных / А.Г.Федоров. - М.: КомпьютерПресс, 2001. – 255 с.

6. Куликовский, Л.Ф. Теоретические основы информационных процессов: учебное пособие / Л.Ф. Куликовский. - М.: ВШ, 2006. – 248 с.

7. Дмитриев, В.И. Прикладная теория информации: учебник / В.И.Дмитриев - М.: ВШ,1989.—320 с.

ДОПОЛНИТЕЛЬНАЯ

8. Левин, М. Д. Методы поиска информации в Интернет / М.Д.Левин. - М.: Солон- Пресс, 2003. – 224 с.

9. Соколов, А.В. Защита информации в распределенных корпоративных сетях и системах / А.В.Соколов. - М.: ДМК-Пресс, 2002. – 655 с.

10. Федоров, Ю.Н. Справочник инженера по АСУТП. Проектирование и разработка: учебно-практическое пособие / Ю.Н. Федоров. – М.: Инфра-Инженерия, 2008. – 928 с.

11. Информационные технологии управления: учебное пособие /под ред. Г.А.Титоренко. - М.: ЮНИТИ-ДАНА, 2011. – 439 с.

12. Шаньгин, В.Ф. Защита информации в компьютерных системах и сетях. / В. Ф. Шаньгин. – М., 2012 – 593 с.

Информация: поиск, анализ, защита. - Минск.: Амалфея, 2002. – 309 с.

13. Запечников, С.В. Криптографические методы защиты информации: учеб. Пособие для академического бакалавриата / С.В. Запечников, О.В. Казарин, А.А. Тарасов. – М., 2016 – 309 с.

14. Барашко, О.Г. Проектирование систем автоматизации: учебно-методическое пособие для студентов специальности 1-53 01 01 «Автоматизация технологических процессов и производств» / О.Г. Барашко. – Минск : БГТУ, 2012.– 221 с.

15. Васильева, И.Н. Криптографические методы защиты информации: учебник и практикум для академического бакалавриата / И. Н. Васильева. – М., 2016 – 349 с.

16. Завгородний, В.И. Комплексная защита информации в компьютерных системах. / В. И. Завгородний – М., 2001– 264 с. Каймин В.А. Информатика. - М.: Инфра-М, 2010.-- 272с.