

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

6. Измерение информации

6.7. Семантический подход к измерению количества информации

Универсальная схема передачи информации



Семантический уровень

Идея:

- Проблемы связаны с формализацией и учетом смысла передаваемой информации.
- Проблемы этого уровня чрезвычайно сложны, так как смысловое содержание информации больше зависит от получателя, чем от семантики сообщения, представленного на каком-либо языке.

На данном уровне:

- Анализируются сведения, которые отражает информация
- Выявляется смысл информации
- Выявляется содержание информации
- Осуществляется обобщение информации

Под тезаурусом (от греческого сокровище) понимается некий обобщенный справочник, определяющий уровень знаний получателя сообщений. Такой справочник включает не только описание понятий, но и связи между ними и т. д.

Идея (Норберт Винер): для понимания и использования информации ее получатель должен обладать определенным запасом знаний

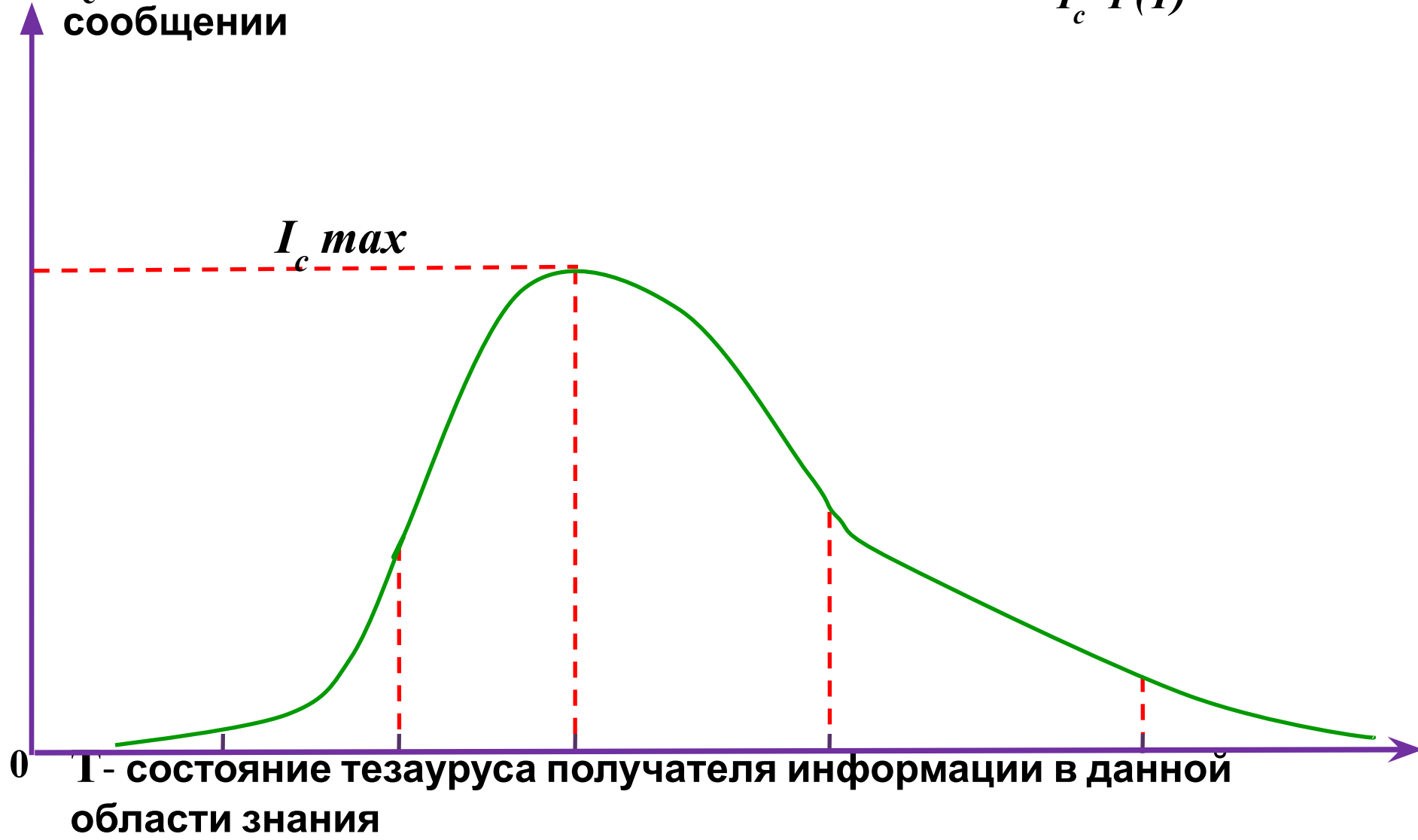
Тезаурусная мера (Ю.А. Шрейдер): основана на учете запаса знаний получателя (приемника) информации

- Сообщения, которые содержат новую для получателя семантическую информацию, изменяют его тезаурус. Под этим понимается не только появление новых понятий, но и установление новых связей между существующими понятиями, ликвидацию устаревших связей или понятий и т.д.
- Если сообщение не вносит ничего нового в тезаурус получателя, то семантическая информация, содержащаяся в сообщении, равна нулю
- Из двух сообщений больше семантической информации будет нести то, которая вносит в тезаурус более существенные

зависимость семантической информации в одном и том же сообщении от тезауруса пользователя

I_c - количество смысловой информации в сообщении

$$I_c = F(T)$$



0 T - состояние тезауруса получателя информации в данной области знания

Измерение количества информации на основе изменения логической вероятности

Под логической вероятностью понимается степень подтверждения той или иной гипотезы

Идея (Р Карнап, Бар-Хиллел):

Посылки:

- Пусть выдвинута некоторая априорная (доопытная) гипотеза и оценена степень подтверждения (логическая вероятность) гипотезы
- Поступающие сообщения могут либо уменьшать степень подтверждения гипотезы, либо оставлять ее неизменной

Вывод:

Количество семантической информации в сообщении будет тем выше, чем больше уменьшится степень подтверждения гипотезы

Связь меры, основанной на изменении логической вероятности и тезаурусной меры

Посылки:

- Очевидно, уменьшение степени подтверждения гипотезы после получения сообщения изменяет априорный запас знаний (тезаурус) получателя информации
- Если же гипотеза построена на эмпирических данных, полностью подтверждаемых сообщением, то логическая вероятность гипотезы не изменяется и сообщение не несет для получателя никакой новой информации

Вывод:

Мера, основанная на изменении логической вероятности, смыкается с тезаурусной, т.к.

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

6. Измерение информации

6.8. Прагматический подход к измерению количества информации

Прагматический подход к измерению количества информации: общие положения

Прагматическая мера определяет ценность информации, ее полезность для достижения цели управления

Изучение отношений между знаками и их потребителями с точки зрения использования получаемой информации и влияния знаков на поведение систем составляет основу прагматической теории информации

Для всех подходов здесь характерно стремление связать понятие прагматической информации с целью, целенаправленным поведением и выдвинуть те или иные количественные меры

Мера А.А. Харкевича

Идея:

Мера ценности информации связывается с вероятностью достижения цели при получении этой информации.

Мера ценности информации определяется как изменение вероятности достижения цели после получения сообщения

$$I_{ц} = \log_a p_1 - \log_a p_0 = \log_a \frac{p_1}{p_0}$$

$$p_1 > p_0$$

$$p_1 = p_0$$

$$p_1 < p_0$$

$$I_{ц} > 0$$

$$I_{ц} = 0$$

$$I_{ц} < 0$$

Подход М.М. Бонгарда

Идея:

- Вводится понятие «полезная информация»
- Тем самым сообщение связывается с тем, какую задачу решает получатель, что он знает до прихода сообщения и как его истолковывает

Этот подход имеет вероятностно-алгебраическую сущность и носит более общий характер, чем подход, предложенный А.А. Харкевичем

Подход Д. Харраха

Цель:

показать, как символическая логика и теория семантической информации могут быть использованы для анализа аспектов человеческой коммуникации

Создана:

модель того, как разумный получатель оценивает последовательность сообщений на основе определенных семантических и прагматических свойств *«логическая модель коммуникаций»*

Предлагается:

обеспечить получателя *«программой обработки сообщений»*, с помощью которой извлекается из получаемых сообщений *«годная к употреблению сумма сообщений»*

Именно к этому результату переработки сообщений, а не к сообщениям в их первоначальной форме могут быть применены количественные меры информации

Созданная Харрахом логическая модель коммуникации служит тем языковым каркасом, в рамках которого программа может

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

7. Кодирование информации

Любое сообщение, подлежащее передаче по каналу связи, сохранению на запоминающем устройстве или переработке, должно быть предварительно закодировано.

Кодирование: основные термины

Кодирование можно определить как процесс представления информации в виде некоторых символов и их последовательностей (**кодовых комбинаций**), причем эти символы могут быть в свою очередь представлены (перекодированы) в виде совокупности физических сигналов той или иной природы

Вообще **кодированием** называется отображение состояния одной физической системы **X** с помощью состояния некоторой другой системы **Y**

Кодирование: основные термины

Пусть имеется некоторая система X , которая случайным образом может принимать одно из своих возможных состояний

$$x_1, x_2, \dots, x_n$$

Требуется закодировать ее с помощью другой системы Y , возможные состояния которой

$$y_1, y_2, \dots, y_m$$
$$m \geq n$$

в этом случае каждое состояние системы X можно закодировать при помощи одного состояния (буквы, символа, сигнала и т.д.) системы Y

$$m < n$$

в этом случае состояния системы X придется кодировать при помощи определенных комбинаций (последовательностей) состояний системы Y

Выбор таких комбинаций и
установление соответствия между
передаваемым сообщением и
этим комбинациями и
называется *кодированием* в узком
смысле этого слова

Кодирование: основные термины

Согласно строгой терминологии

кодированием называется отображение произвольного множества A в множество конечных последовательностей (слов) в некотором алфавите B , а **декодированием** — обратное отображение

Отображающее множество B , включающее в себя множество знаков (символов) и слов (кодовых комбинаций), составленных из этих знаков по определенным правилам и предназначенных для однозначного отображения множества A называется **кодом**

Кодирование: основные термины

Конечное множество знаков (букв, цифр и др. символов), применяемых в том или ином языке, называется **алфавитом** данного языка, а количество различных символов алфавита называют **объемом алфавита**

Таким образом, при рассмотрении вопросов кодирования можно говорить об **алфавите**
кода

Требования, предъявляемые к системе кодирования

1 **взаимная однозначность преобразований** отображаемого множества в отображающее множество при кодировании и обратного преобразования при декодировании

экономичность кодирования, обеспечиваемая оптимизацией средней длины кодовой комбинации, что обеспечивает снижение времени, необходимого для передачи и обработки информации, экономию носителей информации

помехоустойчивость, т.е. возможность обнаружения и исправления ошибок в кодовых комбинациях

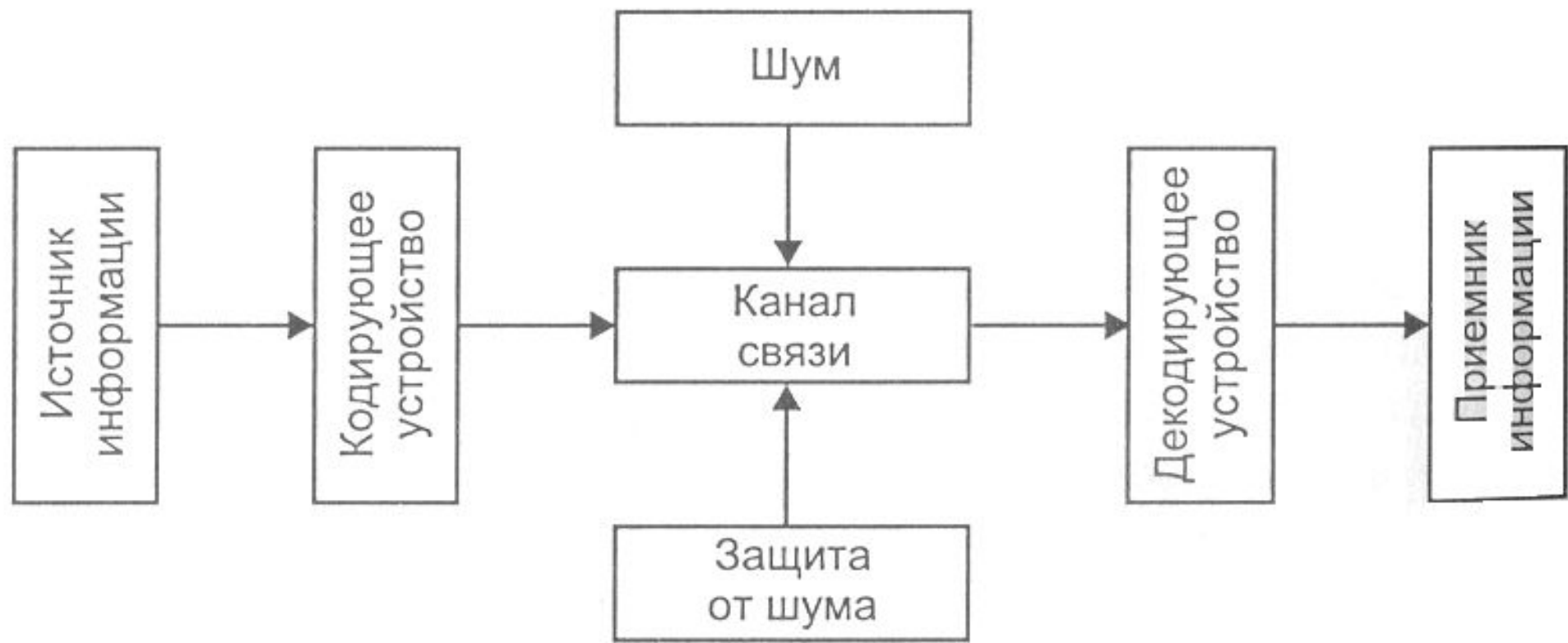
Основные классы кодов

- **равномерные коды**
- **неравномерные коды**
- **неизбыточные**
- **избыточные**
 - **обнаруживающие коды**
 - **корректирующие коды**
 - **систематическими**
 - **несистематическими**

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

7. Кодирование информации

7.1. Помехоустойчивое кодирование



МЕШАЮЩИЕ ВЛИЯНИЯ В КАНАЛАХ СВЯЗИ



Виды линий связи

- проводные линии связи
- радиотракты-радиорелейные прямой видимости
- тропосферные
- космические (через ИСЗ и с дальними космическими кораблями)
- магистральные коротковолновые линии связи
- линии радиосвязи с наземными подвижными объектами
- телефонные каналы
- внутриаппаратные тракты магнитной записи-считывания и шины информационного обмена в компьютерах

Помехи в каналах связи возникают, прежде всего, по техническим причинам: плохое качество линий связи, незащищенность друг от друга различных потоков информации, передаваемой по одним и тем же каналам и т.д.

Передача сообщений при наличии помех является серьезной теоретической и практической задачей. Ее значимость возрастает в связи с повсеместным внедрением компьютерных телекоммуникаций, в которых помехи неизбежны

Основные проблемы при работе с кодированной информацией

- установления самого факта того, что произошло искажение информации
- выяснения того, в каком конкретно месте передаваемого текста это произошло
- исправления ошибки, хотя бы с некоторой степенью достоверности

Устойчивость языка к помехам

В СЛОВОХ

ВСО ГЛОСНОО ЗОМОНОНО БОКВОЙ О

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

7.1. Помехоустойчивое кодирование

7.1.1. Общее представление об избыточности

Коэффициент избыточности сообщения

Коэффициент избыточности r сообщения A

$$r = (I_{max} - I) / I_{max}$$

I - количество информации в сообщении A ;

I_{max} — максимально возможное количество информации в сообщении той же длины, что и A

Наличие избыточности позволяет ставить вопрос о сжатии информации без ее потери в передаваемых сообщениях.

Естественный язык, как способ кодирования. Общее представление об избыточности

Имеется алфавит объемом a символов.

Определить, какое количество N различных слов длиной n символов можно составить в этом алфавите?

$$N = a^n$$

Обычно принимают, что объем русского алфавита равен 32 символа.

Естественный язык, как способ кодирования

$$N = a^n$$

$$n=1$$

32 однобуквенных слова

$$n=2$$

1024 двухбуквенных слова

$$n=3$$

32 768 трехбуквенных слов

$$n=4$$

более миллиона четырехбуквенных слов

$$n=5$$

более 30 миллионов пятибуквенных слов

$$n=6$$

более миллиарда шестибуквенных слов

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ. ИЗМЕРЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

7.1. Помехоустойчивое кодирование

**7.1.1. Контролируемая избыточность
основа помехоустойчивого
кодирования**

**Впервые теоретическое исследование
эффективного кодирования предпринял Клод
Шеннон.**

**Шенноном была разработана специальная
теория кодирования, дающая методы борьбы с
шумом**

**Одна из важных идей этой теории состоит в
том, что передаваемый по линии связи код
должен быть избыточным**

**За счет этого потеря какой-то части
информации при передаче может быть
компенсирована**

**Однако нельзя делать избыточность слишком
большой. Это приведёт к задержкам и**

пропускная способность канала - это максимальное число переданных двоичных единиц (бит) в единицу времени при сколь угодно малой вероятности ошибок.

Реально получаемое число передаваемых бит в единицу времени называют **скоростью передачи**.

При неограниченно малой вероятности ошибок скорость передачи всегда меньше пропускной способности.

В канале с ошибками максимальное значение скорости получают путем использования **помехоустойчивого кодирования**, что требует введения **избыточности** в передаваемый сигнал.

Если код согласован с каналом, введенная избыточность становится оправданной. Если код не согласован с каналом, ошибки могут быть не только не исправлены, но и размножены кодом.

Для согласования кода с каналом связи необходимо иметь максимальный объем сведений о возможных мешающих влияниях в каналах.

Теорема Шеннона

с каждым каналом связано измеряемое в битах в секунду и называемое пропускной способностью канала число C

Если требуемая от системы связи скорость передачи информации R (измеряемая в битах в секунду) меньше C , то, используя коды, контролирующие ошибки, для данного канала можно построить такую систему связи, что вероятность ошибки на выходе будет сколь угодно мала

Выводы из теоремы Шеннона

*построение слишком хороших каналов является
расточительством; экономически выгоднее
использовать кодирование*

*мощность сигнала, шум в канале и полоса
частот ограничивают лишь скорость передачи,
а не ее точность*

Общие свойства систем кодирования

Использование избыточности. Закодированные цифровые сообщения всегда содержат дополнительные, или избыточные, символы. Эти символы используют для того, чтобы подчеркнуть индивидуальность каждого сообщения. Их всегда выбирают так, чтобы сделать маловероятной потерю сообщением его индивидуальности из-за искажения при воздействии помех достаточно большого числа символов.

Усреднение шума. Эффект усреднения достигается за счет того, что избыточные символы зависят от нескольких информационных символов.

Асинхронное и синхронное кодирование

- В *асинхронном режиме* применяют коды, в которых явно выделены границы каждого символа (байта) специальными стартовым и стоповым символами.
- В *синхронном режиме* синхронизм поддерживается во время передачи всего информационного блока без обрамления каждого байта.

Способы контроля правильности передачи данных

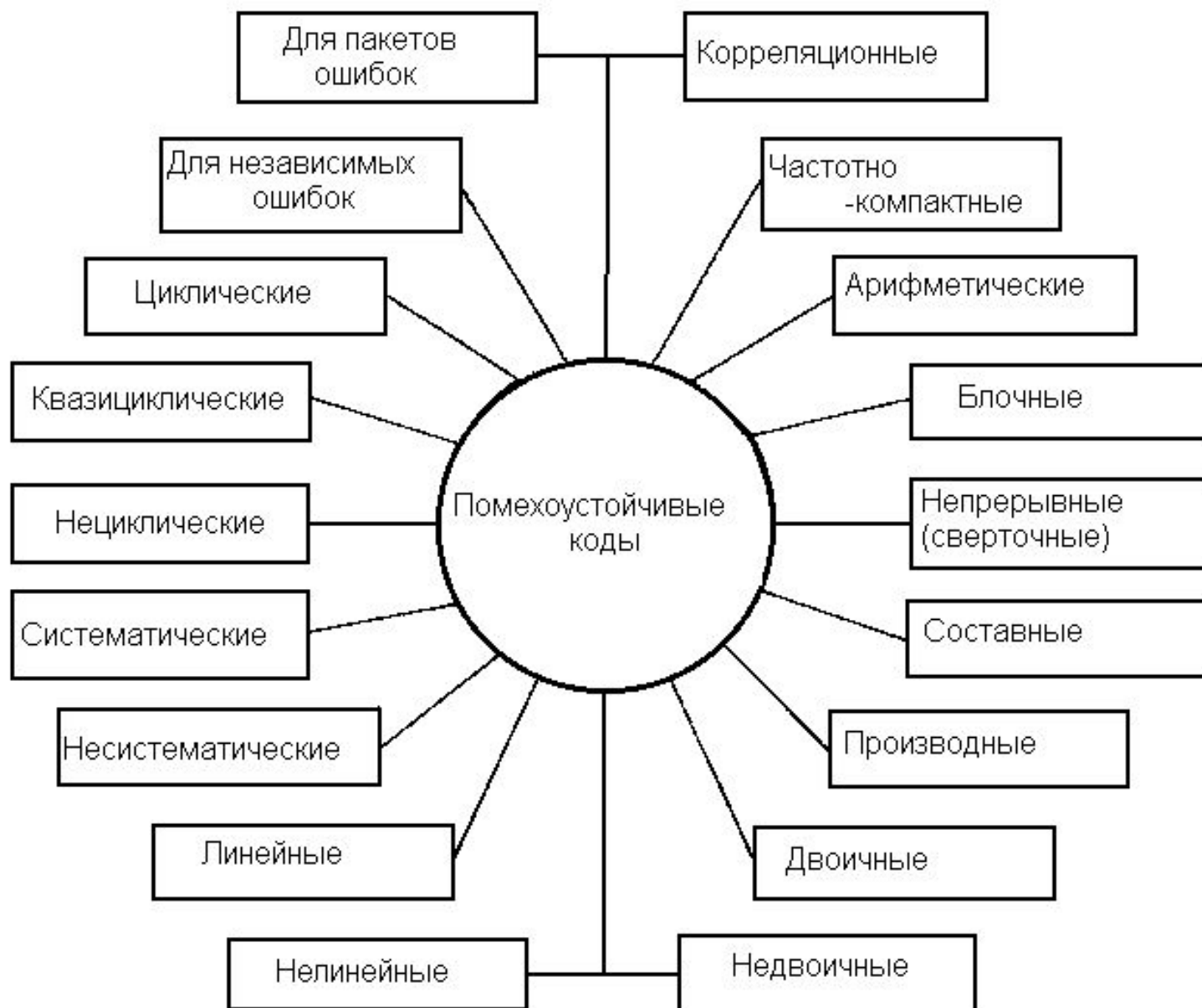
- Различают коды, обнаруживающие ошибки, и корректирующие коды, которые дополнительно к обнаружению еще и исправляют ошибки

- 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.

Краткая классификация помехоустойчивых кодов



Информационные процессы и информационные технологии.

1. Понятие информационного процесса и информационной процедуры

