БЕЛКООПСОЮЗ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ»

Кафедра информационно-вычислительных систем

ТЕОРИЯ ИНФОРМАЦИИ

Практикум
для студентов специальности 1-26 03 01
«Управление информационными ресурсами»

Гомель 2008

УДК 621.391
ББК 32.81
Т 33

Автор-составитель В. В. Бондарева, канд. техн. наук, доцент

Рецензенты: С. С. Стариков, программист I категории
КТСУП «Отель»;
С. М. Мовшович, канд. техн. наук, доцент,
зав. кафедрой информационно-вычислительных систем
Белорусского торгово-экономического университета
потребительской кооперации

Рекомендован к изданию научно-методическим советом учреждения образования «Белорусский торгово-экономический университет потребительской кооперации». Протокол № 1 от 10 октября 2006 г.

Т 33 **Теория** информации : практикум для студентов специальности 1-26 03 01 «Управление информационными ресурсами» / авт.-сост. В. В. Бондарева. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2008. – 44 с.
ISBN 978-985-461-542-4

УДК 621.391
ББК 32.81

ISBN 978-985-461-542-4

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Практикум подготовлен для проведения практических занятий по курсу «Теория информации» для студентов дневной и заочной форм обучения специальности 1-26 03 01 «Управление информационными ресурсами».

Теория информации является математической теорией с высокой степенью общности. Она основывается на теории случайных событий, для описания которых применяются понятия «вероятность» и «энтропия». В самой теории вводится понятие «информация» и устанавливается ее мера – бит.

Примеры использования теории информации можно найти в информатике, технике, психологии, биологии, физике, педагогике, лингвистике и т. д. Однако теория информации применима для решения практических задач в той мере, в какой описываемые материальные системы или процессы соответствуют исходным положениям теории.

Математическое понятие информации связано с возможностью ее количественного измерения. Количественная мера информации не привязывается к ее смысловой основе. Таким образом, теория информации применима для решения лишь тех практических задач, в которых допустимо игнорирование смысловой стороны информации.

Практикум состоит из пяти тем: «Исходные понятия теории информации», «Количественная оценка информации», «Кодирование символьной информации», «Представление и обработка чисел в компьютере», «Передача информации». По каждой теме представлены основные формулы, теоретические сведения, вопросы и задачи. Теоретический материал сопровождается схемами и таблицами.

Числовые данные в задачах в основном носят условный характер. При решении задач студентам не предлагается использовать персональный компьютер.

1. ИСХОДНЫЕ ПОНЯТИЯ ТЕОРИИ ИНФОРМАЦИИ

Основные теоретические сведения

Предмет теории информации

Теория информации как самостоятельная дисциплина возникла в ходе решения следующей задачи: обеспечить надежную и эффективную передачу информации от источника к приемнику при условии, что передаче препятствуют помехи.

Информация (лат. informatio) – сведения, разъяснения, изложение. Это отражение внешнего мира с помощью знаков и сигналов. Информация всегда связана с материальным носителем – материальным объектом или средой, которые служат для представления или передачи информации.

Сигнал – изменение характеристики носителя, которое используется для представления информации. Для передачи информации применяется ряд следующих друг за другом сигналов.

От источника к приемнику информация передается в виде сообщений. Сообщением называется последовательность сигналов. Оно служит переносчиком информации, а информация является содержанием сообщения.

Соответствие между сообщением и содержащейся в нем информацией называется правилом интерпретации сообщения. Это соответствие может быть однозначным и неоднозначным.

С передачей информации связана еще одна пара понятий – источник и приемник информации.

Источник информации – это субъект или объект, порождающий информацию и представляющий ее в виде сообщения.

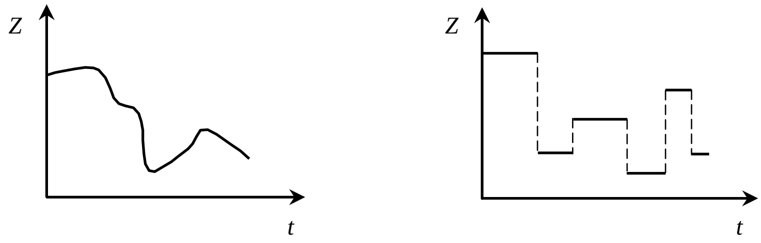
Приемник информации – это субъект или объект, принимающий сообщение и способный правильно его интерпретировать.

Совокупность технических средств, используемых для передачи сообщений от источника к приемнику информации, называется системой связи.

Формы представления информации

Выделяют два типа сигналов: непрерывные и дискретные (рис. 1). Сигнал называется непрерывным (аналоговым), если его параметр может принимать любое значение в пределах некоторого интервала. Сигнал называется дискретным, если его параметр может принимать конечное число значений в пределах некоторого интервала.

Обозначим значение параметра сигнала буквой Z, а время – t.



а) б)

Рис. 1. **Непрерывные (а) и дискретные (б) сигналы**

Принципиальным различием непрерывных и дискретных сигналов является то, что дискретные сигналы можно обозначить, т. е. приписать каждому значению сигнала знак (жест, рисунок, букву), который будет отличать данный сигнал от другого.

Знак – это элемент некоторого конечного множества отличных друг от друга сущностей. Вся совокупность знаков, используемых для представления дискретной информации, называется набором знаков.

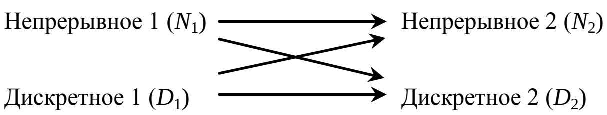
Алфавитом называется набор знаков, в котором установлен порядок их следования. Порядок между знаками устанавливается с помощью отношения «больше – меньше»: для двух знаков α и β принимается, что $\alpha < \beta$, если порядковый номер у α в алфавите меньше, чем у β.

При передаче сообщения параметр сигнала должен меняться. Минимальное количество различных его значений равно двум. Следовательно, такой алфавит содержит минимум два знака и называется двоичным. Верхней границы числа знаков в алфавите не существует.

Знаки, используемые для обозначения фонем человеческого языка, называются буквами, а их совокупность – алфавитом языка. Если знаку приписано содержание, то он будет называться символом.

Преобразование сообщений

Поскольку существуют два типа сообщений, между ними возможны четыре варианта преобразований:



На практике применяются все четыре вида преобразований:

• *Преобразование типа* $N_1 \to N_2$. Такое преобразование всегда сопровождается частичной потерей информации. Потери связаны с внешними помехами и помехами, которые производит само информационное техническое устройство.

• *Преобразования типа* $N \to D$ *и* $D \to N$. Перевод $N \to D$ означает замену описывающей его непрерывной функции $Z(t)$ на некотором отрезке $[t_1, t_2]$ конечным множеством $\{Z_i, t_i\}$ ($i = 0...n$, где n – количество точек разбиения временного интервала).

Преобразование $N \to D$ называется дискретизацией непрерывного сигнала и производится с помощью двух операций: развертки по времени и квантования по величине сигнала (рис. 2).

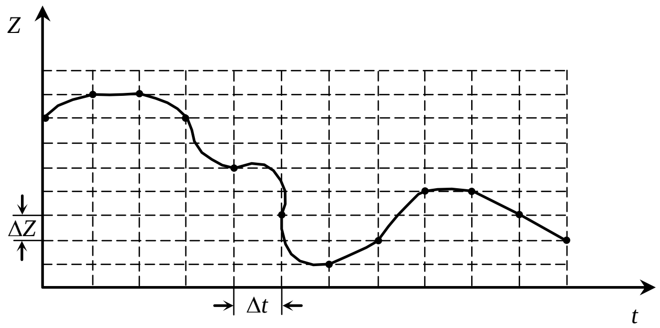


Рис. 2. **Дискретизация непрерывного сигнала**

Развертка по времени состоит в том, что наблюдение за значением величины Z производится не непрерывно, а лишь в определенные моменты времени с интервалом Δt:

$$\Delta t = \frac{t_n - t_0}{n}.$$

Квантование по величине – это отображение вещественных значений параметра сигнала в конечное множество чисел, кратных некоторой постоянной величине – шагу квантования ΔZ.

Теорема Котельникова. Непрерывный сигнал можно полностью отобразить и точно воссоздать по последовательности измерений или отсчетов величины этого сигнала через одинаковые интервалы времени, меньшие или равные половине периода максимальной частоты, имеющейся в сигнале.

По теореме Котельникова преобразования $N \to D$ и $D \to N$, могут осуществляться без потери информации.

Согласно теореме определяющим является значение верхней границы частоты сигнала – ν_m. Смысл теоремы в том, что дискретизация не приведет к потере информации и по дискретным сигналам можно будет полностью восстановить исходный аналоговый сигнал, если развертка по времени выполнена в соответствии со следующим соотношением:

$$\Delta t \le \frac{1}{2\nu_m}.$$

Шаг квантования ΔZ определяется чувствительностью приемного устройства.

• *Преобразование типа* $D_1 \to D_2$. Такое преобразование заключается в переходе от сигналов одного алфавита к сигналам другого алфавита и носит название «перекодировка». Оно может осуществляться без потерь информации.

Вопросы и задания

1. Что является предметом теории информации? Каковы разделы теории информации?
2. Приведите примеры терминов, имеющих несколько трактовок в различных науках, технике, быту.
3. В каких формах может быть представлена информация?
4. Приведите примеры информации:
• в неживой природе;
• в биологических системах;
• в технических устройствах;
• в жизни общества.
5. Приведите примеры процессов, используемых для передачи информации, и связанных с ними сигналов.
6. Являются ли информацией для Вас сведения, содержащиеся в библиотеке затонувшей Атлантиды или в библиотеке Конгресса США?
7. Являются ли информацией нерасшифрованные космические послания?
8. Получаете ли Вы информацию при повторном прочтении книги, учебника?
9. Является ли вакуум носителем информации?
10. В чем различия между понятиями «событие», «сообщение», «информация»?
11. Приведите примеры неоднозначного и однозначного соответствий между сообщением и содержащейся в нем информацией.
12. В чем состоит различие понятий «приемник сообщения» и «приемник информации»?
13. Что такое сигнал? Какие существуют виды сигналов?
14. Приведите примеры непрерывных сигналов.
15. Приведите примеры дискретных сигналов.
16. Органы чувств человека ориентированы на восприятие аналоговой информации. Означает ли это, что человек не может воспринимать информацию дискретную?
17. Во всех примерах, рассмотренных в начале темы, выделите носитель информации и определите, является ли сигнал, передающий эту информацию, дискретным или непрерывным.
18. Приведите примеры процессов, используемых для передачи информации с указанием ее носителя. Какого типа сигнал передает эту информацию?
19. Какой смысл имеет понятие «система связи»?
20. Приведите как минимум пять разных примеров алфавитов. Из них только один может содержать знаки-буквы какой-нибудь известной письменности.
21. Приведите примеры знаков-символов. Могут ли символы образовывать алфавит?
22. В шестнадцатеричной системе счисления используются знаки A, B, C, D, E и F. Следует ли эти знаки считать символами?
23. В чем заключается смысл и значение теоремы отсчетов?
24. Какое количество отсчетов за 1 с необходимо производить цифровому звукозаписывающему устройству, если требуется обеспечить качество записи:
• телефона (4 кГц);
• лазерного диска (20 кГц)?
25. Как следует понимать термины «оцифровка изображения» и «оцифровка звука»? Какими устройствами производятся данные операции?

26. Приведите примеры преобразований типа $D_1 \to D_2$, при которых информация, содержащаяся в исходном сообщении, может не сохраняться.

27. Что такое квантование непрерывных сигналов? С какой целью и погрешностью выполняется квантование? Приведите пример квантования сигнала.

28. Какими достоинствами и недостатками обладают способы преобразования непрерывных сигналов в дискретные путем дискретизации их во времени и квантования по величине сигнала?

29. Каким образом выбрать шаг квантования непрерывного сигнала по уровню, чтобы удовлетворить требованиям заданной точности и помехоустойчивости передачи?

30. Каковы виды квантования? В чем их отличие и каковы цели квантования того или иного вида?

31. В чем заключаются достоинства и недостатки систем, использующих непрерывные сигналы (сообщения)?

32. Почему для представления дискретных сообщений в качестве базового выбирается двоичный алфавит?

33. Почему компьютер является универсальным устройством по обработке информации?

34. В чем состоит и как проявляется несимметричность непрерывной и дискретной форм представления информации?

2. КОЛИЧЕСТВЕННАЯ ОЦЕНКА ИНФОРМАЦИИ

Основные теоретические сведения

Понятие энтропии

Энтропия является мерой неопределенности опыта, в котором проявляются случайные события, и равна средней неопределенности всех возможных его исходов. Обозначим эту величину H.

Энтропия позволяет сравнивать количества информации, приносимые различными сигналами. За единицу измерения принимается неопределенность, содержащаяся в опыте, имеющем лишь два равновероятных исхода, которые можно обозначить как *истину* и *ложь*.

Единица измерения энтропии называется «бит».

В опыте α с n исходами каждый исход случаен и вносит свой вклад в неопределенность всего опыта. Неопределенность, вносимая каждым из исходов, равна

$$H = -p \cdot \log_2 p\,,$$

где p – вероятность отдельного исхода.

Если опыт α имеет n равновероятных исходов, то энтропия такого опыта равна

$$H(\alpha) = \log_2 n\,.$$

Если опыт α имеет n неравновероятных исходов A_1, A_2, ..., A_n, то энтропия такого опыта равна

$$H(\alpha) = -\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i)\,,$$

где $p(A_i)$ – вероятность отдельного A_i исхода.

Опыт $\alpha\wedge\beta$ относится к сложным в том случае, если опыты не являются независимыми, т. е. если на исход β оказывает влияние результат опыта α.

Энтропия сложного опыта $\alpha\wedge\beta$ равна

$$H(\alpha\wedge\beta) = H(\alpha) + H_\alpha(\beta),$$

где $H(\alpha)$ – энтропия опыта α;

$H_\alpha(\beta) = \sum_{i=1}^{n} p(A_i) \cdot H_{A_i}(\beta)$ – средняя условная энтропия опыта β при условии, что в опыте α реализовался исход A_i;

$H_{A_i}(\beta) = -\sum_{i=1}^{n} p_{Ai}(B_j) \cdot \log_2 p_{A_i}(B_j)$ – средняя условная энтропия опыта β при условии выполнения опыта α.

Энтропия и информация

Предшествующий опыт α может уменьшить количество исходов и неопределенность последующего опыта β. Разность $H(\alpha)$ и $H_\alpha(\beta)$ показывает, какие новые сведения относительно β получаем, произведя опыт α. Эта величина называется информацией (I) относительно опыта β, содержащейся в опыте α:

$$I(\alpha, \beta) = H_\alpha(\beta).$$

Это выражение дает возможность *численного измерения количества информации*.

Поскольку единицей измерения энтропии является бит, то в этих же единицах может быть измерено количество информации.

Энтропия опыта равна той информации, которую получаем в результате его осуществления.

Информация, содержащаяся в опыте α с n неравновероятными исходами $A_1, A_2, ..., A_n$, равна

$$I = -\sum_{i=1}^{n} p(A_i) \cdot \log_2 p(A_i).$$

Количество информации численно равно числу вопросов с равновероятными бинарными вариантами ответов, которые необходимо задать, чтобы полностью снять неопределенность задачи.

Когда все n исходов опыта равновероятны, т. е. $p(A_i) = \dfrac{1}{n}$, тогда информация, содержащая в опыте α, равна

$$I = \sum_{i=1}^{n} \frac{1}{n} \cdot \log_2 n = \log_2 n.$$

Эта формула была выведена в 1928 г. американским инженером Р. Хартли и носит его имя. Ее смысл в том, что, если некоторое множество содержит n элементов, то для однозначной идентификации одного элемента среди прочих требуется количество информации, равное $\log_2 n$.

Частным случаем применения последней формулы является ситуация, когда $n = 2^k$. Подставляя это значение в формулу, получим

$$I = k.$$

Информация и алфавит

При передаче сообщения возникает проблема распознавания знака, т. е. по полученным сигналам необходимо установить исходную последовательность знаков первичного алфавита.

В устной речи это достигается использованием звуков. В письменности это достигается различным начертанием букв и дальнейшим анализом написанного. Узнавание знака требует получения некоторой порции информации. Можно связать эту информацию с самим знаком и считать, что знак несет в себе некоторое количество информации. Оценим это количество. Начнем с самого грубого приближения и назовем его нулевым.

Нулевое приближение. Пусть появление всех знаков алфавита в сообщении равновероятно. Тогда для английского алфавита число знаков $n_e = 27$ (с учетом пробела), для русского алфавита число знаков $n_r = 34$. Из формулы Хартли находим:

$$I_0^{(e)} = \log_2 27 = 4{,}755 \text{ бита};$$

$$I_0^{(r)} = \log_2 34 = 5{,}087 \text{ бита}.$$

Первое приближение. Вероятность появления различных букв в тексте различна. Рассмотрим таблицу средних частот букв русского алфавита, в который включен знак «пробел» для разделения слов, с учетом неразличимости букв «е» и «ё», а также «ь» и «ъ» (так принято в телеграфном кодировании). Получим алфавит из 32 знаков с вероятностями их появления в русских текстах (табл. 1).

Таблица 1. **Средние частоты букв и знаков русского алфавита**

Буква, знак	Относительная частота	Буква, знак	Относительная частота
Пробел	0,174	я	0,018
о	0,090	ы	0,016
е, ё	0,072	з	0,016
а	0,062	ь, ъ	0,014
и	0,062	б	0,014
т	0,053	г	0,013
н	0,053	ч	0,012
с	0,045	й	0,010
р	0,040	х	0,009
в	0,038	ж	0,007
л	0,035	ю	0,006
к	0,028	ш	0,006
м	0,026	ц	0,004
д	0,025	щ	0,003
п	0,023	э	0,003
у	0,021	ф	0,002

Среднее количество информации, приходящейся на один знак, равно

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

где p_i – вероятность i-го знака данного алфавита из n знаков.

Это знаменитая формула К. Шеннона, с работы которого «Математическая теория связи» (1948) принято начинать отсчет возраста информатики как самостоятельной науки.

Сообщения, в которых вероятность появления каждого отдельного знака не меняется со временем, называются шенноновскими, а порождающий их отправитель – шенноновским источником.

Если сообщение является шенноновским, то набор знаков (алфавит) и связанная с каждым знаком информация известны заранее. В этом случае интерпретация сообщения сводится к задаче распознавания знака. Теория информации строится именно для шенноновских сообщений, поэтому в дальнейшем будем рассматривать только такие сообщения.

Применение формулы Шеннона дает значение средней информации на знак алфавита:

$$I_1^{(r)} = 4{,}36 \text{ бит}, \quad I_1^{(e)} = 4{,}04 \text{ бит}.$$

Второе приближение. Значение средней информации на букву может быть уменьшено с учетом связей (корреляцией) между буквами в словах. В словах буквы появляются не в любых сочетаниях. Учет в английских словах двухбуквенных сочетаний понижает среднюю информацию на знак до следующих значений:

$$I_2^{(e)} = 3{,}32 \text{ бит}, \quad I_2^{(r)} = 3{,}52 \text{ бит}.$$

Учет трехбуквенных сочетаний дает следующую среднюю информацию:

$$I_3^{(e)} = 3{,}10 \text{ бит}, \quad I_3^{(r)} = 3{,}01 \text{ бит}.$$

Последовательность I_0, I_1, I_2 является убывающей в любом языке. Шеннон ввел величину, которую назвал относительной избыточностью языка.

Относительная избыточность языка (R) равна

$$R = 1 - \frac{I_\infty}{I_0} \,,$$

где I_∞ – предельная информация на знак в данном языке;

I_0 – наибольшая информация, которая может содержаться в знаке данного алфавита.

Вопросы и задания

1. Что такое энтропия сообщений, источника сообщений? Укажите, как ее рассчитать:
- в случае равновероятных дискретных сообщений;

• для неравновероятных дискретных сообщений.

2. Почему в определении энтропии как меры неопределенности выбрана логарифмическая зависимость между H и n? Почему выбран $\log_2$?

3. Укажите энтропию следующих опытов:

• бросок монеты;

• бросок игральной кости;

• вытаскивание наугад одной игральной карты из 36;

• бросок двух игральных костей.

4. Двигатель троллейбуса может работать в одном из пяти режимов. Вероятность того, что он работает в первом режиме $p_1 = 0{,}08$, во втором – $p_2 = 0{,}12$, в третьем – $p_3 = 0{,}15$, в четвертом – $p_4 = 0{,}28$, в пятом – $p_5 = 0{,}37$. Найдите энтропию множества возможных режимов работы двигателя.

5. В первой урне имеются 7 белых, 5 черных и 2 синих шара. Во второй урне – 4 белых, 6 черных и 2 синих шара. Наудачу вынимают один шар. Для какой из урн исход более определенный?

6. Алфавит русского языка содержит 34 буквы (с пробелом), английского – 27. Если считать появление всех букв в тексте одинаковым, то как соотносятся неопределенности, связанные с угадыванием случайно выбранной буквы текста?

7. Опыт имеет два исхода. Докажите, что энтропия такого опыта максимальна, если обе вероятности исходов будут равны 0,5.

8. На соревнованиях по стрельбе на мишенях нанесены области с очками 0, 2, 6, 10. Стрелки А и В сделали по 100 выстрелов и показали результаты, представленные в табл. 2.

Таблица 2. **Результаты соревнований по стрелкам**

Стрелок	Число попаданий в области с очками			
	0	**2**	**6**	**10**
А	5	20	65	10
В	10	10	60	20

Определите, с результатом выстрела которого из стрелков (А или В) связана большая неопределенность.

9. Имеются два ящика, в каждом из которых по 12 шаров. В первом ящике находятся 3 белых, 3 черных и 6 красных шаров, во втором – по 4 шара каждого цвета. Опыты заключаются в вытаскивании по одному шару из каждого ящика. Шары возвращаются в ящик после извлечения. Что можно сказать относительно неопределенностей исходов этих опытов?

10. Опыты α и β состоят в последовательном извлечении без возврата двух шаров из ящика, в котором изначально находились n белых шаров и m черных. Найдите $H(\alpha)$ и $H_\alpha(\beta)$.

11. Имеются три тела с одинаковыми внешними размерами, но с разными массами m_1, m_2 и m_3. Необходимо определить энтропию, связанную с нахождением наиболее тяжелого из них, если сравнивать массу тел можно только попарно.

12. Источник генерирует знак z_1 с вероятностью 0,8 и z_2 с вероятностью 0,2. Какова энтропия источника?

13. Какими единицами измеряется количество информации? Приведите примеры.

14. Пусть у кого-то из ваших знакомых родился ребенок, и Вы спрашиваете: «Кто родился: мальчик или девочка?». Какое количество информации содержится в ответе?

15. При угадывании результата броска игральной кости задается вопрос: «Выпало 6?» Какое количество информации содержит ответ?

16. Укажите, какое количество информации связано с исходом следующих опытов:

• бросок игральной кости;

• бросок двух монет;

• вытаскивание наугад одной игральной карты из 36;

• бросок двух игральных костей.

17. Сообщение записано в виде десятичного числа из пяти цифр, причем предполагается, что все цифры равновероятностны и независимы. Какое количество информации несет это сообщение? Во сколько раз меньшее количество информации содержало бы сообщение, состоящее из пяти двоичных цифр?

18. Игра «Угадай-ка-4». Некто задумал целое число в интервале от 0 до 3. Предполагается, что вероятности быть задуманными у всех чисел одинаковы. Наш опыт состоит в угадывании этого числа. На наши вопросы можно отвечать лишь «да» или «нет». Какое количество информации нужно получить, чтобы узнать задуманное число, т. е. полностью снять начальную неопределенность? Как правильно построить процесс угадывания?

19. Случайным образом вынимается карта из колоды в 32 карты. Как построить угадывание?

20. Мы отгадываем задуманное кем-то двузначное число. Какое количество информации требуется для отгадывания всего числа? Какова оптимальная последовательность вопросов при отгадывании? Каково их минимальное число?

Изменится ли требуемое количество информации, если будем отгадывать не все число сразу, а по очереди: сначала 1-ю цифру числа, затем – 2-ю? Одинакова ли информация, необходимая для отгадывания 1-й и 2-й цифр?

21. Сколько вопросов в игре «Угадай-ка-7» (угадать число от 0 до 6) нужно задать для определения числа?

22. Одновременно проводится шесть игр «Угадай-ка-7» (угадать число от 0 до 6). Сколько вопросов необходимо задать, если необходимо отгадать одну из возможных комбинаций?

23. Используя понятия энтропии и информации по Шеннону, оцените, сколько вопросов надо задать, чтобы отгадать задуманное собеседником натуральное число, не превосходящее N, если он дает только ответы «да» или «нет».

24. В некоторой местности имеются две близкорасположенные деревни А и В. Известно, что жители деревни А всегда говорят правду, а жители деревни В всегда лгут. Известно также, что жители обеих деревень любят ходить друг к другу в гости, поэтому в каждой из деревень можно встретить жителя соседней деревни. Путешественник, сбившись ночью с пути, оказался в одной из двух деревень и, заговорив с первым встречным, захотел выяснить, в какой деревне он находится и откуда его собеседник. Какое минимальное количество вопросов с бинарными ответами требуется задать путешественнику?

25. Ответьте на вопрос задания 24 при условии, что помимо деревень А и В имеется деревня С, жители которой дают по очереди то правдивые, то ложные ответы, причем неизвестно, с какого они начинают.

26. Вопрос имеет два варианта ответа. Возможно ли, чтобы с каждым из ответов было связано различное количество информации?

27. Как оценить количество информации в непрерывных сообщениях?

28. Возможно ли, чтобы бинарный ответ содержал меньше 1 бита информации?

29. Какое количество информации содержит каждый из ответов на вопрос, если всего их три и все они равновероятны? Если равновероятных ответов n, определите, какое количество информации содержит каждый из ответов.

30. Источник порождает множество шестизнаковых сообщений, каждое из которых содержит один знак «*», два знака «%» и три знака «!» Какое количество информации содержится в каждом (одном) из таких сообщений?

31. С какой из букв русского алфавита («а» или «б») связано больше информации? Найдите эту информацию.

32. Средняя длина слова в русском языке составляет 5,3 буквы, в английском – 4,5. Найдите вероятности появления в соответствующих текстах пробелов. Какое количество информации связано с пробелом в обоих языках?

33. Дайте объяснение тому, что количество информации на знак алфавита выражается нецелым числом.

34. Что такое «шенноновские сообщения»? Почему теория информации имеет дело именно с такими сообщениями?

35. Почему используется «избыточный» язык?

36. Одинакова ли избыточность литературных и деловых текстов? Поясните свой ответ.

37. В Петрозаводске 280 000 жителей. Какое минимальное количество вопросов, требующих ответа «да» или «нет», необходимо, чтобы найти одного жителя.

38. Орудие стреляет по удаленной цели. При каждом выстреле цель поражается с вероятностью $p = 0{,}1$. Разведка может только один раз проверить, поражается ли цель хоть один раз. Через какое количество выстрелов k следует провести проверку, чтобы она дала максимальное количество информации?

39. В лотерее разыгрывается N билетов, из них k выигрышных. Студент купил M билетов и после розыгрыша сообщил Вам, что выиграл (но, возможно, и не на один билет). Какое количество информации Вы получили?

40. Бросаются одновременно две игральные кости. Определите количество информации, содержащееся в сообщении о том, что произведение чисел выпавших очков четно.

41. Два стрелка, для которых вероятности попадания в мишень равны соответственно 0,6 и 0,7, производят по одному выстрелу. В результате оказалось, что мишень поражена. Какое количество информации содержится в этом сообщении?

42. Имеется n одинаковых монет, одна из которых легче. Сколько взвешиваний на чашечных весах необходимо и достаточно, чтобы ее найти?

43. Имеется 12 монет одного достоинства; 11 из них имеют одинаковую массу, а одна (фальшивая) отличается по массе от остальных (причем неизвестно, легче она или тяжелее настоящих). Каково наименьшее число взвешиваний на чашечных весах без гирь, которое позволяет обнаружить фальшивую монету и выяснить, легче она, чем остальные монеты, или тяжелее?

44. Каково количество информации (в битах) содержится в 5-буквенных словах русского и латинского алфавитов (например, английского языка), если предположить, что все буквы (символы алфавита) равновероятны?

45. Символы азбуки Морзе могут появиться в сообщении со следующими вероятностями: для точки – 0,51, для тире – 0,31, для промежутка между буквами – 0,12, для промежутка между словами – 0,06. Определите среднее количество информации в сообщении из 500 символов данного алфавита, считая, что связь между последовательными символами отсутствует.

46. Известно, что из 100 изготовленных деталей в среднем 10 деталей имеют дефекты. Для выявления брака используется метод отбраковки, дающий всегда отрицательный эффект, если деталь изготовлена с браком. Если брак отсутствует, то деталь признается годной лишь в 80% случаев. Какое количество информации о качестве детали можно получить в среднем по результату такого метода отбраковки?

47. Радиотехническое устройство состоит из пяти блоков (А, Б, В, Г, Д). Блок А в среднем выходит из строя 1 раз в 100 дней, блок Б – 1 раз в 25 дней, блок В – 1 раз в 5 дней, блок Г – 1 раз в 4 дня и блок Д – 1 раз в 2 дня. Контрольный прибор позволяет за одно измерение проверить работоспособность в целом любой комбинации блоков. Как нужно проводить контроль, чтобы затратить на поиски неисправного блока в среднем минимальное количество проверок? Найдите это среднее значение.

48. Прямоугольник разделен на три продольные и пять поперечных полос. Найдите количество информации для следующих сообщений:

- точка находится в прямоугольнике 34 (в третьей продольной полосе и четвертой поперечной);
- точка находится во второй продольной полосе;
- точка находится в первой поперечной полосе;
- точка находится в первой продольной и третьей поперечной полосах.

3. КОДИРОВАНИЕ СИМВОЛЬНОЙ ИНФОРМАЦИИ

Основные теоретические сведения

Постановка задачи кодирования

Источник представляет информацию в форме дискретного сообщения, используя для этого алфавит, который называется первичным. Сообщение попадает в устройство, преобразующее и представляющее его в другом алфавите, который называется вторичным.

Код – 1) правило, описывающее соответствие знаков первичного алфавита или их сочетаний знакам или их сочетаниям вторичного алфавита; 2) набор знаков вторичного алфавита, используемый для представления знаков или их сочетаний первичного алфавита.

Кодирование – перевод информации, представленной сообщением в первичном алфавите, в последовательность кодов. *Декодирование* – восстановление информации в первичном алфавите по полученной последовательности кодов.

Операции кодирования и декодирования называются обратимыми**,** если их последовательное применение обеспечивает возврат к исходной информации без каких-либо ее потерь. Рассмотрим обратимое кодирование.

Пусть первичный алфавит А состоит из N знаков со средней информацией на знак $I^{(A)}$, а вторичный алфавит В – из M знаков со средней информацией на знак $I^{(B)}$. Пусть исходное сообщение, представленное в первичном алфавите, содержит n знаков, а закодированное сообщение – m знаков. Условие обратимости кодирования следующее:

$$nI^{(A)} \le mI^{(B)} \quad \text{или} \quad I^{(A)} \le \frac{m}{n} \cdot I^{(B)}.$$

Операция обратимого кодирования может увеличить количество информации в сообщении, но не может его уменьшить. Отношение m/n характеризует среднее число знаков вторичного алфавита, которое приходится использовать для кодирования одного знака первичного алфавита. Оно называется длиной кода. Обозначим его K(А, В). Следовательно,

$$K(A,B) \ge \frac{I^{(A)}}{I^{(B)}}.$$

Обычно $N > M$ и $I(A) > I(B)$, откуда K(А, В) > 1, т. е. один знак первичного алфавита представляется несколькими знаками вторичного. Способов построения кодов при фиксированных алфавитах А и В существует множество. Возникает проблема выбора оптимального кода, который при передаче информации позволяет затратить на передачу сообщения меньше времени.

Минимально возможное значение средней длины кода равно

$$K^{\min}(A,B) = \frac{I^{(A)}}{I^{(B)}}. \qquad (3.1)$$

Это выражение устанавливает нижний предел длины кода, но неясно, какое возможно приближение K(A, B) к $K^{\min}$(A, B). По этой причине для теории кодирования и теории связи важнейшее значение имеют две теоремы, доказанные Шенноном. Первая теорема затрагивает ситуацию с кодированием при отсутствии помех, искажающих сообщение. Вторая теорема относится к реальным линиям связи с помехами.

Первая теорема Шеннона. При отсутствии помех всегда возможен такой вариант кодирования сообщения, при котором среднее число знаков кода, приходящихся на один знак первичного алфавита, будет сколь угодно близко к отношению средней информации на знак первичного и вторичного алфавитов.

В ситуации, рассмотренной К. Шенноном, при кодировании сообщения в первичном алфавите учитывается различная вероятность появления знаков, но их зависимости не отслеживаются. Источники подобных сообщений называются *источниками без памяти*. Если при этом обеспечена равная вероятность появления знаков вторичного алфавита, то, как следует из выражения 3.1, для минимальной средней длины кода оказывается справедливым следующее соотношение:

$$K^{\min}(A,B) = \frac{I_1^{(A)}}{\log_2 M}. \qquad (3.2)$$

Наиболее важной для практики оказывается ситуация, когда $M = 2$, т. е. используется два типа сигналов. Такое кодирование называется двоичным. Технически это наиболее просто реализуемый вариант, например, существование напряжения в проводе – импульс (1) или отсутствие – пауза (0). Удобство двоичных кодов заключается и в том, что при равных длительностях и вероятностях каждый элементарный сигнал несет в себе 1 бит информации ($\log_2 2 = 1$). Тогда из формулы 3.2 следует

$$K^{\min}(A, 2) = I_1^{(A)}.$$

Превышение K(A, B) над $K^{\min}$(A, B) называется относительной избыточностью кода $Q(A, B)$ и рассчитывается по формуле

$$Q(A,B) = \frac{K(A,B) - K^{\min}(A,B)}{K^{\min}(A,B)} = \frac{K(A,B)\cdot I^{(B)}}{I^{(A)}} - 1.$$

В частном случае относительная избыточность кода для двоичных сообщений источника без памяти при кодировании знаками равной вероятности определяется следующим образом:

$$Q(A,2) = \frac{K(A,2)}{I_1^{(A)}} - 1.$$

Способы построения двоичных кодов

В случае неравномерного кодирования необходимо *построить такую схему кодирования, в которой суммарная длительность кодов при передаче данного сообщения была бы наименьшей*. Длительность сообщения будет меньше, если знакам первичного алфавита, которые встречаются в сообщении чаще, присвоить меньшие по длине коды, а тем, которые встречаются реже, – более длинные.

Параллельно должна решаться проблема различимости кодов. Если бы код был равномерным, приемное устройство при декодировании просто отсчитывало бы заданное число элементарных сигналов (например, 5) и интерпретировало их в соответствии с кодовой таблицей. При использовании неравномерного кодирования возможны два подхода к обеспечению различимости кодов:

- используется комбинация элементарных сигналов, которая интерпретируется декодером как разделитель знаков;
- применяются префиксные коды без разделителя.

Правила построения неравномерного кода с разделителем

Пусть разделителем отдельных кодов букв будет последовательность 00 (признак конца знака), а разделителем слов – последовательность 000 (пробел). Код строится по следующим правилам:

- коды всех букв будут заканчиваться 00;
- коды букв не должны содержать двух и более нулей подряд в середине;
- код буквы (кроме пробела) всегда должен начинаться с 1;

• разделителю слов 000 всегда предшествует признак конца знака, при этом реализуется последовательность 00000.

Существует возможность построения двоичного неравномерного кода без разделителя, т. е. варианта кодирования сообщения, при котором последующее выделение из него каждого отдельного знака (декодирование) оказывается однозначным без специальных указателей разделения знаков. Наиболее простыми и употребимыми кодами такого типа являются так называемые префиксные коды.

Код называется *префиксным*, если он удовлетворяет *условию Фано*: неравномерный код может быть однозначно декодирован, если никакой из кодов не совпадает с началом какого-либо иного более длинного кода.

Правила построения префиксного кода Шеннона-Фано

Рассмотрим схему построения префиксного кода Шеннона-Фано на примере. Пусть имеется первичный алфавит А, состоящий из шести знаков $a_1...a_6$ с вероятностями появления в сообщении 0,3; 0,2; 0,2; 0,15; 0,1; 0,05.

Код строится по следующим правилам:

1. Расположим знаки в порядке убывания вероятностей.

2. Разделим знаки на две группы так, чтобы суммы вероятностей в каждой из них были приблизительно равными. В примере в первую группу попадут знаки a_1 и a_2 с суммой вероятностей 0,5, во вторую – остальные четыре знака с суммой вероятностей также 0,5.

3. Первой группе присвоим первый знак кода 0. Второй группе – первый знак кода 1.

4. Продолжим деление каждой из групп на подгруппы по этой же схеме, т. е. так, чтобы суммы вероятностей на каждом шаге в соседних подгруппах были бы возможно более близкими.

Результат представлен в табл. 3.

Таблица 3. **Построение кода Шеннона-Фано**

Знак	Вероятность (p_i)	Разряды кода				Код
		1	**2**	**3**	**4**	
a_1	0,3	0	0			00
a_2	0,2	0	1			01
a_3	0,2	1	0			10
a_4	0,15	1	1	0		110
a_5	0,1	1	1	1	0	1110
a_6	0,05	1	1	1	1	1111

Код удовлетворяет условию Фано и, следовательно, является префиксным.

Правила построения префиксного кода Хаффмана

Рассмотрим построение префиксного кода Хаффмана на примере. Пусть имеется первичный алфавит А, состоящий из шести знаков $a_1...a_6$ с вероятностями появления в сообщении 0,3; 0,2; 0,2; 0,15; 0,1; 0,05.

Код строится по следующим правилам:

1. Расположим знаки в порядке убывания вероятностей (табл. 4).

Таблица 4. **Построение кода Хаффмана (шаги 1–5)**

Знак	Вероятности				
	Исходный алфавит (A)	Промежуточные алфавиты			
		$A^{(1)}$	$A^{(2)}$	$A^{(3)}$	$A^{(4)}$
a_1	0,3	0,3	0,3	0,4	0,6
a_2	0,2	0,2	0,3	0,3	0,4
a_3	0,2	0,2	0,2	0,3	
a_4	0,15	0,15	0,2		
a_5	0,1	0,15			
a_6	0,05				

2. Создадим новый алфавит $A^{(1)}$, объединив два знака с наименьшими вероятностями a_5 и a_6 и заменив их одним знаком $a^{(1)}$ с вероятностью, равной сумме вероятностей знаков a_5 и a_6.

3. Остальные знаки включим в новый алфавит без изменения. Общее число знаков в новом алфавите будет на 1 меньше, чем в исходном.

4. Переупорядочим знаки по убыванию вероятностей.

5. Продолжим создавать новые алфавиты, пока в последнем не останется два знака.

Всю процедуру построения представим в виде табл. 4.

6. Проведем кодирование в обратном направлении. Двум знакам последнего алфавита присвоим коды: верхнему знаку – 0, а нижнему – 1. В алфавите $A^{(4)}$ знак $a_1^{(4)}$, имеющий вероятность 0,6, получит код 0, а знак $a_2^{(4)}$ с вероятностью 0,4 – код 1.

В алфавите $A^{(3)}$ знак $a_1^{(3)}$ получает от знака $a_2^{(4)}$ его вероятность 0,4 и код 1; коды знаков $a_2^{(3)}$ и $a_3^{(3)}$, происходящие от знака $a_1^{(4)}$ с вероятностью 0,6, будут уже двузначными: их первой цифрой станет код их «родителя» – 0, а вторая цифра, как и раньше, у верхнего – 0, у нижнего – 1; таким образом, знак $a_2^{(3)}$ будет иметь код 00, а знак $a_3^{(3)}$ – код 01 и т. д.

Процедура кодирования представлена в табл. 5.

Таблица 5. **Построение кода Хаффмана (шаг 6)**

Знак	Вероятности									
	Исходный алфавит (A)		Промежуточные алфавиты							
			$A^{(1)}$		$A^{(2)}$		$A^{(3)}$		$A^{(4)}$	
a_1	0,3	00	0,3	00	0,3	00	0,4	1	0,6	0
a_2	0,2	10	0,2	10	0,3	01	0,3	00	0,4	1
a_3	0,2	11	0,2	11	0,2	10	0,3	01		
a_4	0,15	010	0,15	010	0,2	11				
a_5	0,1	0110	0,15	011						
a_6	0,05	0111								

Код удовлетворяет условию Фано и не требует разделителя.

Код Хаффмана обладает более высокой эффективностью по сравнению с кодом Шеннона-Фано, если сравнить избыточности кодов для естественного языка. Код Хаффмана является самым экономичным из всех возможных, т. е. *ни для какого метода алфавитного кодирования длина кода не может оказаться меньше, чем код Хаффмана*.

Вопросы и задания

1. Какое из двух преобразований $y = \lg x$ и $y = \cos x$ является взаимно однозначным кодированием применительно к положительным числам?

2. Требуется декодировать следующее сообщение в соответствии с табл. 6:

00100010000111010101110000110.

Таблица 6. **Таблица префиксных кодов**

а	л	м	р	у	ы
10	010	00	11	0110	0111

3. Приведите примеры обратимого и необратимого кодирования.

4. В чем смысл первой теоремы Шеннона для кодирования?

5. Какие источники сообщений называются источниками без памяти? Приведите пример источника с памятью и источника без памяти.

6. Первичный алфавит содержит 8 знаков с вероятностями (табл. 7).

Таблица 7. **Вероятности знаков первичного алфавита**

Знак	Пробел	?	&	*	+	%	#	!
Вероятность	0,25	0,18	0,15	0,12	0,1	0,08	0,07	0,05

В соответствии с правилами предложите вариант неравномерного алфавитного двоичного кода с разделителем знаков, а также постройте коды Шеннона-Фано и Хаффмана. Сравните их избыточности.

7. Почему в 1 байте содержится 8 бит?

8. Укажите, какое количество книг объемом 200 страниц может поместиться:

- на дискете емкостью 1,44 Мб;
- в ОЗУ компьютера емкостью 32 Мб;
- на CD-диске емкостью 650 Мб;
- на жестком магнитном диске винчестера емкостью 8,4 Гб.

9. Почему в компьютерных устройствах используется байтовое кодирование?

10. Считая алфавит цифр самостоятельным, а появление различных цифр равновероятным, найдите избыточность кода Морзе для цифрового алфавита (табл. 8).

Таблица 8. **Код Морзе для цифр**

Цифра	Код	Цифра	Код
0	– – – – –	5	· · · · ·
1	· – – – –	6	– · · · ·
2	· · – – –	7	– – · · ·
3	· · · – –	8	– – – · ·
4	· · · · –	9	– – – – ·

11. Пусть первичный алфавит состоит из знаков a и b с вероятностями 0,75 и 0,25. Сравните избыточность кода Хаффмана при алфавитном и блочном двухбуквенном кодировании.

12. В лексиконе Эллочки Щукиной из романа Ильфа и Петрова «12 стульев» было 17 словосочетаний: «Хо-хо!», «Ого!», «Блеск!», «Шутишь, парниша», «У вас вся спина белая» и пр.

Определите длину кода при равномерном двоичном кодировании. Предложите вариант равномерного кодирования данного словарного запаса.

13. Источник генерирует знаки z_1 с вероятностью 0,8 и z_2 с вероятностью 0,2. Постройте эффективные коды Шеннона-Фано и Хаффмана для последовательности из трех знаков $z_i z_j z_k$. Каково среднее число символов на знак?

14. Первичный алфавит содержит 7 знаков с вероятностями 0,3; 0,22; 0,18; 0,12, 0,08; 0,06 и 0,04. В соответствии с правилами постройте коды Шеннона-Фано и Хаффмана. Сравните их избыточности.

15. Известно, что энтропия, приходящаяся на одну букву русского алфавита, составляет приблизительно 1,2 бита. Определите минимальное количество десятичных символов, необходимых для передачи информации, содержащейся в телеграмме из 100 букв.

4. ПРЕДСТАВЛЕНИЕ И ОБРАБОТКА ЧИСЕЛ В КОМПЬЮТЕРЕ

Основные теоретические сведения

Системы счисления

Любое число имеет *значение* и *форму представления*. Значение числа задает его отношение к значениям других чисел (>, <, =). Форма представления определяет порядок записи числа с помощью предназначенных для этого знаков. *Значение числа не зависит от способа его представления*. Способ представления числа определяется *системой счисления*.

Система счисления – это правило записи чисел с помощью заданного набора специальных знаков – цифр.

Системы счисления делятся следующим образом: *унарная*, *непозиционные* и *позиционные*.

В настоящее время для представления чисел применяют, в основном, позиционные системы счисления.

Позиционными называются системы счисления, в которых значение каждой цифры в изображении числа определяется ее позицией в ряду других цифр. Наиболее распространенной является система счисления, в которой для записи чисел используется 10 цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8 и 9. Число представляет собой краткую запись многочлена, в который входят степени некоторого другого числа – *основания системы счисления*.

Вещественное число можно представить в виде суммы целого числа и правильной дроби.

Пусть p – основание системы счисления. Тогда целое число Z может быть представлено в следующем виде:

$$Z_p = a_{k-1} \cdot p^{k-1} + a_{k-2} \cdot p^{k-2} + \ldots + a_1 \cdot p^1 + a_0 \cdot p^0, \qquad (4.1)$$

где k – общее число цифр ($k > 0$, целое);
a_j – целые числа, $0 \le a_i \le p - 1$.

Правильная дробь $0,Y$ может быть представлена в следующем виде:

$$0,Y_p = b_1 \cdot p^{-1} + b_2 \cdot p^{-2} + \ldots + b_n \cdot p^{-n}, \qquad (4.2)$$

где n – общее число цифр ($n > 0$, целое);
b_j – целые числа, $0 \le b_j \le p - 1$.

Перевод целых чисел из одной системы счисления в другую

Перевод числа из системы p в q обозначим $Z_p \rightarrow Z_q$. Его возможно произвести при любых q и p, но он затруднен тем, что придется выполнять операции по правилам арифметики недесятичных систем счисления. Поэтому удобными являются преобразования $Z_p \rightarrow Z_r \rightarrow Z_q$. Z_r – система счисления с основанием r, для которой арифметические операции выполнить легко. Такими основаниями являются $r = 1$ и $r = 10$. Рассмотрим систему счисления с основанием 10.

В преобразовании $Z_p \rightarrow Z_{10} \rightarrow Z_q$ первая и вторая части не связаны друг с другом, поэтому рассмотрим их по отдельности.

Алгоритм перевода $Z_{10} \rightarrow Z_q$:

1. Целочисленно разделить исходное число Z_{10} на основание новой системы счисления q и найти остаток от деления. Это будет цифра 0-го разряда числа Z_q.
2. Частное от деления снова целочисленно разделить на q с выделением остатка; процедуру продолжать до тех пор, пока частное от деления не окажется меньше q.
3. Образовавшиеся остатки от деления, поставленные в порядке, обратном порядку их получения, и представляют Z_q.

Алгоритм перевода $Z_p \rightarrow Z_{10}$:

1. Число Z_p представить в виде многочлена по формуле 4.1.
2. Выполнить все операции по правилам десятичной арифметики.

Перевод дробных чисел из одной системы счисления в другую

Дробь в системе счисления p обозначим $0,Y_p$, дробь в системе q – $0,Y_q$, преобразование обозначим $0,Y_p \rightarrow 0,Y_q$. Осуществлять преобразование удобно через промежуточный переход к десятичной системе, т. е. $0,Y_p \rightarrow 0,Y_{10} \rightarrow 0,Y_q$. Это, в свою очередь, разбивает задачу на две составляющие: преобразования $0,Y_p \rightarrow 0,Y_{10}$ и $0,Y_{10} \rightarrow 0,Y_q$, каждое из которых может рассматриваться независимо.

Алгоритм перевода $0,Y_{10} \rightarrow 0,Y_q$:

1. Умножить исходную дробь в десятичной системе счисления на q, выделить целую часть (она будет первой цифрой новой дроби). Отбросить целую часть.
2. Для оставшейся дробной части операцию умножения с выделением целой и дробных частей повторять, пока в дробной части не окажется 0 или не будет достигнута желаемая точность конечного числа. Появляющиеся при этом целые числа будут цифрами новой дроби.
3. Записать дробь в виде последовательности чисел после нуля с разделителем в порядке их появления в пунктах 1 и 2.

Алгоритм перевода $0,Y_p \rightarrow 0,Y_{10}$:

1. Число $0,Y_p$ представить в виде многочлена по формуле 4.2.
2. Выполнить все операции по правилам десятичной арифметики.

Обработка целых чисел без знака

Правила сложения двоичных чисел:

$0 + 0 = 0$; $\quad 0 + 1 = 1$;
$1 + 0 = 1$; $\quad 1 + 1 = 10$.

Правила умножения двоичных чисел:

$0 \cdot 0 = 0$; $\quad 0 \cdot 1 = 0$;
$1 \cdot 0 = 0$; $\quad 1 \cdot 1 = 1$.

Вопросы и задания

1. Почему унарную систему счисления нельзя считать позиционной с основанием 1?
2. Запишите следующие числа в десятичной системе счисления:

а) 100011_2; д) $11011{,}01_2$; и) $1051{,}3_6$;
б) $1021{,}1_3$; е) 1332_4; к) $101{,}3_9$;
в) 1011_7; ж) $106{,}2_7$; л) $2fa_{16}$;
г) $35{,}6_8$; з) $15a_{16}$; м) $19{,}a_{16}$.

3. Переведите десятичные числа в заданные системы счисления:

а) $36 = \_\_\_\__2$; д) $996 = \_\_\_\__6$; и) $397 = \_\_\_\__{16}$;
б) $197 = \_\_\_\__3$; е) $899 = \_\_\_\__7$; к) $8769 = \_\_\_\__{16}$;
в) $984 = \_\_\_\__4$; ж) $98 = \_\_\_\__8$; л) $5397 = \_\_\_\__{16}$;
г) $63 = \_\_\_\__5$; з) $769 = \_\_\_\__{16}$; м) $6997 = \_\_\_\_\__2$.

4. Представьте числа в системе счисления, указанной в правой части:

а) $142_5 = \_\_\__{10}$; д) $142_7 = \_\_\__5$; и) $316_8 = \_\_\__2$;
б) $142_{10} = \_\_\__5$; е) $100110010_2 = \_\_\__4$; к) $316_8 = \_\_\__{10}$;
в) $142_8 = \_\_\__{10}$; ж) $100110010_2 = \_\_\__{10}$; л) $1212_4 = \_\_\__8$;
г) $142_5 = \_\_\__7$; з) $316_8 = \_\_\__4$; м) $1212_4 = \_\_\__2$.

5. Преобразуйте десятичные числа в двоичные и восьмеричные:

а) 20; д) 65; и) 129;
б) 5; е) 127; к) 1135;
в) 15; ж) 1024; л) 92;
г) 32; з) 2047; м) 109.

6. Преобразуйте двоичные числа в восьмеричные и десятичные:

а) 101; д) 111111; и) 100000;
б) 1001; е) 1100100; к) 111011;
в) 1101; ж) 100100; л) 11001;
г) 100001; з) 101010; м) 10010.

7. Преобразуйте восьмеричные числа в шестнадцатеричные:

а) 102235; г) 77777; ж) 15136;
б) 16; д) 177776; з) 17332;
в) 47; е) 70450; и) 11673.

8. Переведите шестнадцатеричные числа в восьмеричные:

а) 1f; г) abcd; ж) f67a5dc;
б) e2; д) ffff; з) 799a6f3;
в) f1; е) 1ffe; и) d5a92f.

9. Переведите в двоичную систему десятичные числа:

а) 0,625; д) 56,25; и) 0,01;
б) 0,28125; е) 127,75; к) 0,122;
в) 0,0078125; ж) 0,69; л) 0,1207;
г) 0,04525; з) 0,375; м) 345,23.

10. Произведите преобразование $Z_p \to Z_q$ без перехода к промежуточным системам счисления:

а) $120112_3 = \_\_\_\__9$; б) $AAA_{16} = \_\_\_\__4$; в) $12345678_9 = \_\_\_\__3$.

11. Определите десятичное значение числа, представленного в двоичной системе счисления с помощью n единиц.

12. Исследуйте экономичность различных систем счисления для общего количества цифр 24, 60, 120. Сделайте вывод.

13. Выполните арифметические действия в заданных системах счисления:

а) $126_7 + 662_7$;
б) $126_8 + 662_8$;
в) $101101_2 + 11011_2$;
г) $6346_8 \cdot 447_8$;
д) $1044_6 \cdot 305_6$;
е) $1011_2 \cdot 1101_2$;
ж) $6512_7 + 566_7$;
з) $3765_8 + 122_8$;
и) $101_2 + 11_2$;
к) $11_8 \cdot 11_8$;
л) $135_6 \cdot 23_6$;
м) $1111_2 \cdot 101_2$;
н) $a81c_{16} + 9fb6_{16}$;
о) $a1c_{16} \cdot 6_{16}$;
п) $1c_{16} \cdot ab_{16}$.

5. ПЕРЕДАЧА ИНФОРМАЦИИ

Основные теоретические сведения

Общая схема передачи информации

Способов передачи информации существует множество: почта, телефон, радио, телевидение, компьютерные сети и пр. Однако в них можно выделить общие элементы (рис. 3).

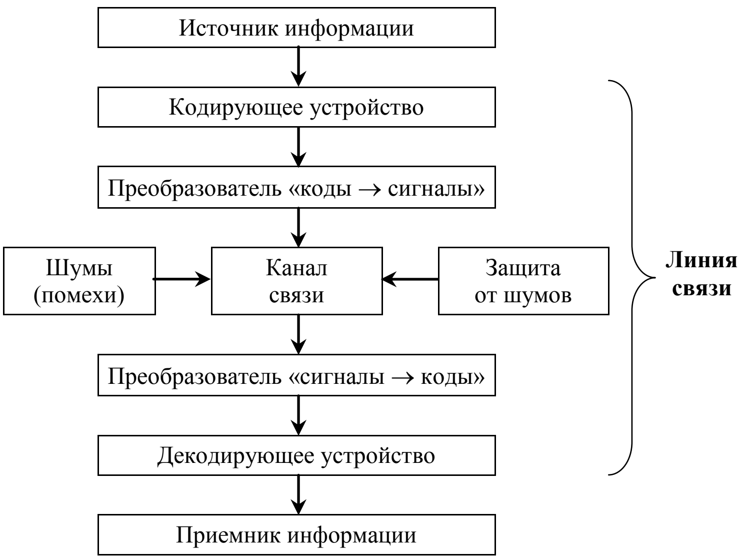


Рис. 3. **Общая схема передачи информации**

Источник информации порождает информацию и для передачи представляет ее в виде сообщения. Для представления информации он использует систему кодирования. *Кодирующее устройство* может являться подсистемой источника или внешним устройством по отношению к источнику информации.

Преобразователь переводит коды в последовательность материальных сигналов, т. е. помещает их на материальный носитель. Преобразователь может быть совмещен с кодирующим устройством, но может быть и самостоятельным элементом линии связи.

После преобразователя сигналы поступают и распространяются по *каналу связи*. Реальный канал связи подвержен внешним воздействиям, а также в нем могут происходить внутренние процессы, в результате которых искажаются передаваемые сигналы и связанное с ними сообщение. Такие воздействия называются шумами (помехами).

После прохождения сообщения по каналу связи сигналы с помощью преобразователя переводятся в последовательность кодов, которые затем декодируются в форму необходимой приемнику информации. На этапе приема преобразователь также может быть совмещен с декодирующим устройством или существовать самостоятельно.

Линия связи объединяет все элементы от источника до приемника информации. Характеристиками линии связи являются скорость передачи сообщения, а также степень искажения сообщения в процессе передачи.

Характеристики канала связи

Рассмотрим каналы связи, передача сообщений по которым осуществляется за счет электрических импульсов.

Ширина полосы пропускания – интервал частот, используемый каналом связи для передачи сигналов.

В данном случае важна не сама ширина полосы пропускания, а максимальное значение частоты (ν_m) из данной полосы, поскольку именно она определяет возможную скорость передачи информации по каналу.

Длительность элементарного импульса определяют исходя из следующих соображений. Если параметр сигнала меняется синусоидально, то за один период колебания T сигнал будет иметь одно максимальное и одно минимальное значения.

Если аппроксимировать синусоиду прямоугольными импульсами и сместить начало отсчета на уровень минимального значения (рис. 4), получится, что сигнал принимает всего два значения: максимальное – *импульс* (обозначим его 1) и минимальное – *пауза* (0).

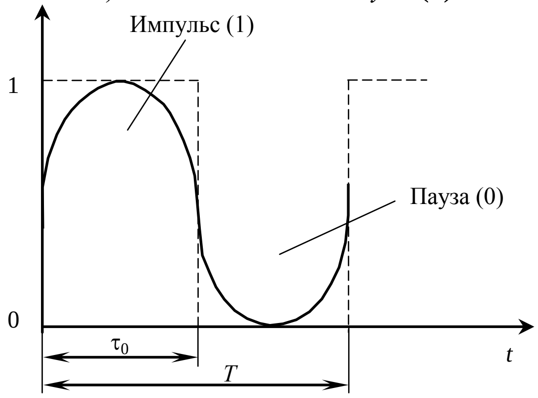


Рис. 4. **Аппроксимация синусоидального параметра сигнала**

Импульс и пауза – элементарные сигналы. При выбранной аппроксимации длительности сигналов (τ_0) одинаковы и равны:

$$\tau_0 = \frac{T}{2} = \frac{1}{2\nu_m}.$$

Если же импульсы порождаются тактовым генератором, имеющим частоту ν_m, то

$$\tau_0 = \frac{1}{\nu_m}.$$

Таким образом, ν_m определяет длительность элементарного сигнала τ_0, используемого для передачи сообщения.

Пропускная способность канала связи. Если с передачей одного импульса связано количество информации I_{imp}, а передается оно за время τ_0, то отношение I_{imp} к τ_0 будет отражать среднее количество информации, передаваемое по каналу за единицу времени. Эта величина является характеристикой канала связи и называется пропускной способностью канала (C), которая рассчитывается по формуле

$$C = \frac{I_{imp}}{\tau_0}.$$

Если I_{imp} выражено в битах, а τ_0 – в секундах, то единицей измерения C будет бит/с. Величину I_{imp} можно установить следующим образом: если первичный алфавит содержит N знаков с вероятностями появления их в сообщении p_i, то по формуле Шеннона можно найти среднюю информацию на знак первичного алфавита I_1, для представления которого используется двоичный код длиной $K^{(2)}$. Тогда

$$I_{imp} = \frac{I_1}{K^{(2)}}.$$

Отсюда получаем формулу для определения пропускной способности канала:

$$C = \frac{I_1}{K^{(2)} \cdot \tau_0}.$$

Скорость передачи информации. Пусть по каналу связи за время t передано количество информации I. Можно ввести величину, характеризующую быстроту передачи информации – скорость передачи информации J:

$$J = \frac{I}{t}.$$

Размерностью J, как и C, является бит/с.

Максимальная скорость передачи информации по каналу связи равна его пропускной способности.

Влияние шумов на пропускную способность канала

Отличие реального канала от идеального состоит в том, что шумы приводят к снижению пропускной способности канала.

В реальном канале из-за шумов при передаче может произойти искажение сигнала. Пусть его вероятность равна p. Тогда вероятность того, что исходный сигнал придет без искажений, равна $1–p$. Следовательно, при распознавании конечного сигнала возникает неопределенность, которая может быть охарактеризована средней энтропией:

$$H = -p \cdot \log_2 p - (1-p) \cdot \log_2(1-p). \quad (5.1)$$

Эта неопределенность приведет к уменьшению количества информации, содержащейся в сигнале, на величину H:

$$(I_{imp})' = I_{imp} - H .$$

Поскольку длительность импульса τ_0 определяется частотой ν_m и не зависит от наличия шумов, пропускная способность реального канала C_R оказывается меньше, чем аналогичного идеального C:

$$C_R = \frac{(I_{imp})'}{\tau_0} = \frac{I_{imp} + p \cdot \log_2 p + (1-p) \cdot \log_2(1-p)}{\tau_0} \le C .$$

Обеспечение надежности передачи и хранения информации

Все реальные каналы связи подвержены воздействию помех. Шеннон доказал теоретическую возможность передачи сообщения без потерь информации по реальным каналам, если при этом выполнен ряд условий.

Вторая теорема Шеннона. При передаче информации по каналу с шумом всегда имеется способ кодирования, при котором сообщение будет передаваться со сколь угодно высокой достоверностью, если скорость передачи не превышает пропускной способности канала.

Решение проблемы состоит в использовании таких методов кодирования информации, которые позволили бы контролировать правильность передачи и при обнаружении ошибки исправлять ее. При этом можно рассмотреть две ситуации:

- кодирование обеспечивает *только установление факта* искажения информации, в этом случае исправление производится путем ее повторной передачи;
- кодирование позволяет *определить и автоматически исправить* ошибку передачи.

Общим условием является использование только равномерных кодов. Надежность обеспечивается тем, что с битами, кодирующими сообщение (информационными битами), передаются дополнительные биты (контрольные), по состоянию которых можно судить о правильности передачи. При *равномерном кодировании* сообщения длина кодовой цепочки k для знака складывается из длины информационной части k_i и числа контрольных битов k_c.

Избыточность сообщения (L) для реального канала определим следующим образом:

$$L = \frac{k}{k_i} = \frac{k_i + k_c}{k_i} = 1 + \frac{k_c}{k_i} .$$

Относительная избыточность сообщения – это характеристика, показывающая, во сколько раз требуется удлинить сообщение, чтобы обеспечить его надежную передачу (хранение). При равной надежности передачи предпочтение должно быть отдано тому способу кодирования, при котором избыточность окажется наименьшей.

Избыточность сообщения характеризует эффективность кодирования при передаче по реальным каналам, но не указывает, каким образом следует осуществлять кодирование, чтобы ошибка могла быть локализована и определена.

Произведем количественные оценки. Шумы в канале связи ведут к частичной потере передаваемой информации на величину возникающей неопределенности, которая при передаче одного бита исходного сообщения может быть вычислена по формуле 5.1.

Для восстановления информационного содержания сообщения следует дополнительно передать количество информации не менее величины ее потерь, т. е. вместо передачи каждого 1 бита информации следует передавать $1 + H$ битов. В этом случае избыточность сообщения составит

$$L_{\min} = \frac{1+H}{1} = 1 - p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p).$$

Избыточность $L_{\min}$ считается минимальной, поскольку при передаче сообщения по каналу, характеризуемому вероятностью искажения p и избыточностью меньше $L_{\min}$, восстановление информации оказывается невозможным.

Рассмотрим метод кодирования, позволяющий определить, в каком бите находится ошибка. Метод был предложен в 1948 г. Р. Хеммингом. Построенные по этому методу коды получили название «коды Хемминга».

Код Хемминга

Идея состоит в добавлении к информационным битам нескольких битов четности, каждый из которых контролирует определенные информационные биты. Если пронумеровать все передаваемые биты, начиная с первого слева направо (информационные биты нумеруются с нулевого и справа налево), то контрольными оказываются биты, номера которых равны степеням числа 2, а все остальные являются информационными. Например, для восьмибитного информационного кода (рис. 5) контрольными окажутся биты с номерами 1, 2, 4 и 8:

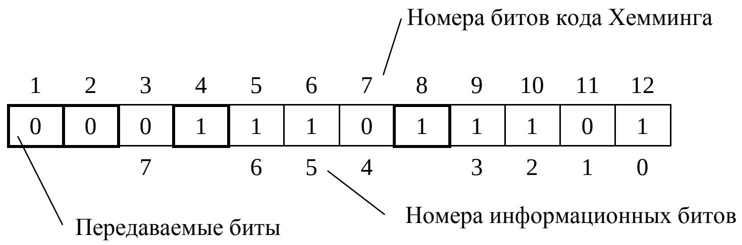


Рис. 5. **Восьмибитный информационный код**

Номера контролируемых битов для каждого проверочного бита приведены в табл. 9. В перечень контролируемых битов входит и тот, в котором располагается проверочный. При этом состояние проверочного бита устанавливается таким образом, чтобы суммарное количество единиц в контролируемых им битах было четным.

Таблица 9. **Номера контролируемых битов для проверочных**

Номера проверочных битов	Номера контролируемых битов											
1	1	3	5	7	9	11	13	15	17	19	21	…
2	2	3	6	7	10	11	14	15	18	19	22	…
4	4	5	6	7	12	13	14	15	20	21	22	…
8	8	9	10	11	12	13	14	15	24	25	26	…
16	16	17	18	19	20	21	22	23	24	25	26	…
32	32	33	34	35	36	37	38	39	40	41	42	…

Для любого номера проверочного бита n, начиная с него, n битов подряд оказываются проверяемыми, затем следует группа из n непроверяемых битов, далее происходит чередование групп.

Номер бита, содержащего ошибку, равен сумме номеров контрольных битов, указавших на ее существование.

Алгоритм проверки и исправления передаваемой последовательности бит в представлении Хемминга:

1. Произвести проверку всех битов четности.
2. Если все биты четности верны, то перейти к пункту 5.
3. Вычислить сумму номеров всех неправильных битов четности.
4. Инвертировать содержимое бита, номер которого равен сумме, найденной в пункте 3.
5. Исключить биты четности, передать правильный информационный код.

Избыточность кодов Хемминга для различных длин передаваемых последовательностей приведена в табл. 10.

Видно, что выгоднее передавать и хранить более длинные последовательности битов. При этом избыточность не должна оказаться меньше $L_{\min}$ для выбранного канала связи.

Таблица 10. **Избыточность кодов Хемминга**

Число информационных битов	Число контрольных битов	Избыточность, *L*
8	4	1,50
16	5	1,31
32	6	1,06

Вопросы и задания

1. Приведите примеры каналов связи, среды их существования, носителей сообщения и процессов, используемых для передачи сообщения.

2. Какими факторами определяется близость реального канала связи к идеальному?

3. Приведите примеры процессов, используемых для передачи информации, и связанных с ними сигналов, кроме указанных в тексте.

4. В информационном канале используется алфавит с четырьмя различными символами. Длительность всех символов одинакова и равна $t = 1$ мкс. Определите пропускную способность канала при отсутствии шумов.

5. Что произойдет при попытке передачи информации со скоростью, превышающей пропускную способность канала связи? Поясните свой ответ.

6. Человек может осмысленно читать со скоростью 15 знаков в секунду. Оцените пропускную способность зрительного канала в данном виде деятельности, если считать появления букв равновероятностными.

7. Оцените пропускную способность слухового канала радиста, принимающего сигналы азбуки Морзе, если известно, что для распознавания одного элементарного сигнала ему требуется 0,2 с.

8. Почему происходит потеря информации при ее передаче по каналу с шумом?

9. Определите, на какую долю снижается пропускная способность канала с шумом по сравнению с идеальным каналом при двоичном кодировании, если вероятность появления ошибки передачи составляет:

- 0,001;
- 0,02;
- 0,1;
- 0,5;
- 0,98.

Поясните полученные результаты.

10. Почему при передаче информации предпочтение отдается равномерному коду?

11. В чем смысловое отличие понятия «избыточность» для идеальных и реальных каналов передачи информации?

12. Почему оказывается невыгодной передача длинных кодовых цепочек с одним проверочным битом?

13. Будет ли установлен факт ошибки передачи, если эта ошибка содержится в самом контрольном бите? Обоснуйте свой ответ.

14. Какое минимальное количество контрольных битов должно передаваться вместе с 16-ю информационными битами для обеспечения восстановимости информации, если вероятность искажения составляет 0,001; 0,02; 0,1; 0,5; 0,98? Какова реальная избыточность сообщения в каждом случае?

15. Получено машинное слово, закодированное с использованием кода Хемминга:

- 101111110001;
- 010001000111;
- 100010111100010110011;
- 101010010110101100001.

Устраните ошибку передачи.

16. В каких ситуациях код Хемминга не позволит локализовать и исправить ошибку передачи?

17. Алфавит обитателей планеты Тау-Кита содержит следующий набор знаков:

∗, ⊥, <, >, =, &, #, ↵, ∇.

Предложите вариант равномерного двоичного кодирования этого алфавита, а также определите избыточность кода при последовательной передаче с одним битом четности.

18. По каналу в одну секунду передается 106 символов (скорость передачи 106 бит/с). Символы 0 и 1 поступают на вход канала с равной вероятностью. Определите пропускную способность канала при следующих условиях:

- символ 1 воспринимается как 1 с вероятностью 0,9 и как 0 с вероятностью 0,1, так же искажается и символ 0;

• в последовательности из четырех символов искажается один символ с вероятностью 0,1.

19. Источник создает последовательность из алфавита, содержащего 16 равновероятных и статистически независимых букв. При передаче по каналу с шумом буквы искажаются так, что четверть всех букв принимается неправильно, причем все ошибки одинаково вероятны. Определите среднюю информацию в принятой букве относительно переданной.

20. Двоичный стирающий канал является одним из наиболее простых типов канала с шумом. В нем переданные символы могут быть «стертыми», но никогда не могут быть приняты ошибочно. Найдите среднее количество информации, переносимое одним символом в таком канале:

• если вероятность стирания равна 0,1 и не зависит от переданного символа;

• если вероятности символов на входе одинаковы.

21. Алфавит состоит из 8 согласных и 8 гласных букв. В сообщении все буквы алфавита равновероятностны и независимы. После прохождения через канал связи согласные всегда принимаются безошибочно, а гласные – только в половине случаев. В половине случаев имеют место ошибки «случайного перепутывания гласных букв». Какое количество информации по Шеннону содержится в принятом символе о переданном символе?

22. Источник дискретных сообщений генерирует независимую последовательность из алфавита, состоящего из 16 равновероятностных символов. При передаче по каналу связи символы искажаются так, что четверть всех символов алфавита принимается неправильно, причем все ошибки равновероятностны. Вычислите шенноновское количество информации в принятом символе о переданном символе.

23. Студент может получить зачет с вероятностью 0,3, не проработав весь учебный материал, и с вероятностью 0,9, проработав его полностью. Какое количество информации о степени проработки учебного материала можно получить по результату сдачи зачета, если в среднем 90% студентов полностью проработали учебный материал?

24. Статистика прогнозирования дождя в городе А представлена в табл. 11.

Таблица 11. **Статистика прогнозирования дождя**

Прогноз	Истинное состояние	
	Дождь	Нет дождя
Дождь	2/16	3/16
Нет дождя	1/16	10/16

Студент заметил, что бюро прогнозов не ошибается лишь в 12 случаях из 16. Если всегда давать прогноз «Нет дождя», то доля правильных решений окажется больше: 13 из 16. Студент предложил этот способ прогнозирования с просьбой о гонораре. Однако начальник бюро прогнозов как специалист по теории информации посчитал проект студента необоснованным. Поясните его решение.

25. На входе канала связи имеется последовательность двоичных независимых равновероятностных символов. Ошибки в канале приводят к изменению значений некоторых символов на обратные, причем вероятность ошибки в k-м символе зависит лишь от наличия ошибки в предыдущем $(k-1)$-м символе: она равна 0,2 при наличии ошибки в $(k-1)$-м символе и 0,05 при ее отсутствии. Найдите среднее количество передаваемой информации в расчете на символ.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Галлагер, Р. Теория информации и надежная связь / Р. Галлагер. – М. : Совет. радио, 1974. – 720 с.

Дмитриев, В. И. Прикладная теория информации : учеб. для студентов вузов по специальности «Автоматизированные системы обработки информации и управления» / В. И. Дмитриев. – М. : Высш. шк., 1989. – 320 с.

Колесник, В. Д. Курс теории информации / В. Д. Колесник, Г. Ш. Полтырев. – М. : Наука, 1982. – 416 с.

Стратонович, Р. Л. Теория информации / Р. Л. Стратонович. – М. : Совет. радио, 1975. – 424 с.

Стариченко, Б. Е. Теоретические основы информатики : учеб. пособие для вузов / Б. Е. Стариченко. – М. : Горячая линия – Телеком, 2004. – 312 с.

Чисар, И. Теория информации: теоремы кодирования для дискретных систем без памяти / И. Чисар, Я. Кернер. – М. : Мир, 1985. – 400 с.

СОДЕРЖАНИЕ

Учебное издание

ТЕОРИЯ ИНФОРМАЦИИ

Практикум
для студентов специальности 1-26 03 01
«Управление информационными ресурсами»

Автор-составитель **Бондарева** Валентина Викторовна

Редактор Е. Г. Привалова
Технический редактор И. А. Козлова
Компьютерная верстка Л. Ф. Кириленкова

Подписано в печать 31.01.08. Бумага типографская № 1.
Формат 60 × 84 $^{1}/_{16}$. Гарнитура Таймс. Ризография.
Усл. печ. л. 2,56. Уч.-изд. л. 2,80. Тираж 150 экз.
Заказ №

Учреждение образования
«Белорусский торгово-экономический
университет потребительской кооперации».
246029, г. Гомель, просп. Октября, 50.
ЛИ № 02330/0056814 от 02.03.2004 г.

Отпечатано в учреждении образования
«Белорусский торгово-экономический
университет потребительской кооперации».
246029, г. Гомель, просп. Октября, 50.