

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
«ЧЕРНІГІВСЬКИЙ КОЛЕГІУМ» імені Т.Г. ШЕВЧЕНКА

Л. А. Матвійчук

ТЕОРЕТИЧНІ ОСНОВИ ІНФОРМАТИКИ

*Рекомендовано вченою радою
Національного університету «Чернігівський колегіум»
імені Т.Г. Шевченка як навчальний посібник
для студентів вищих навчальних закладів*

Чернігів – 2018

УДК 621.391(075.8)
ББК 32.811я73
М54

Рецензенти:

Горошко Ю. В. – доктор педагогічних наук, професор кафедри інформатики і обчислювальної техніки Національного університету «Чернігівський колегіум» імені Т. Г. Шевченка;

Войтович І. С. – доктор педагогічних наук, професор, завідувач кафедри комп'ютерної інженерії та освітніх вимірювань Факультету інформатики НПУ імені М. П. Драгоманова

Матвійчук Л. А.

М 54 Теоретичні основи інформатики : навчальний посібник / Л. А. Матвійчук. – Чернігів : Національний університет «Чернігівський колегіум» імені Т.Г. Шевченка, 2018. – 104 с.

У навчальному посібнику висвітлено основні положення теорії інформації, кодування, теорії автоматів, розпізнавання образів, кібернетики. Наведено питання для самоконтролю, відомості щодо підготовки та виконання самостійної роботи студентів, тести, екзаменаційні питання, термінологічний словник, що сприяє кращому засвоєнню навчального матеріалу.

Посібник укладено відповідно програми дисципліни «Теоретичні основи інформатики», який є нормативною дисципліною навчального плану підготовки спеціалістів, магістрів напрямів: 111 Математика, 014.08 Середня освіта (Фізика), 014.04 Середня освіта (Математика), 014.09 Середня освіта (Інформатика), а також буде корисним для студентів інших спеціальностей, які цікавляться проблемами інформації, інформатики, кодування та інформології в цілому.

УДК 621.391(075.8)
ББК 32.811я73

*Рекомендовано до друку вченою радою
Національного університету «Чернігівський колегіум» імені Т.Г. Шевченка
протокол №__ від _____ 2018 р.*

© Матвійчук Л.А., 2018

ЗМІСТ

Вступ	7
ЗМІСТОВИЙ МОДУЛЬ І. ТЕОРЕТИЧНІ ОСНОВИ ІНФОРМАТИКИ	9
Тема 1. Теорія інформації	9
Питання для самоконтролю	27
Тема 2. Теорія кодування	27
Питання для самоконтролю	53
Тема 3. Теорія автоматів	53
Питання для самоконтролю	59
Тема 4. Теорія розпізнавання образів	60
Питання для самоконтролю	64
Тема 5. Кібернетика	64
Питання для самоконтролю	67
ЗМІСТОВИЙ МОДУЛЬ ІІ. САМОСТІЙНА РОБОТА	68
<i>Вимоги до самостійної роботи</i>	68
Теми самостійної роботи	68
Модульна контрольна робота	84
Екзаменаційні питання	88
ГЛОСАРІЙ	91
Додатки	93
Список використаних джерел	99

Вступ

Стрімкі процеси інформатизації впливають на всі сфери людської діяльності, вимагаючи від спеціалістів все більш досконалішого розуміння актуальних тенденцій, вміння ефективно вирішувати завдання за допомогою обчислювальної техніки.

Інформатика – це технічна наука, яка систематизує процес створення, збереження, накопичування, опрацювання інформації за допомогою електронно-обчислювальної техніки. Отже, інформатика вирішує велику кількість завдань, з якими стикається сьогодні суспільство.

Поняття інформації займає центральне місце в інформатиці, в основу якої входять такі розділи, як: теоретична інформатика, штучний інтелект, програмування, прикладна інформатика, кібернетика тощо.

Кібернетика розглядає інформатику, штучний інтелект та програмування з точки зору здатності сприймати певну інформацію, зберігати, передавати. Термін «кібернетика» ще 2500 років тому запропонував давньогрецький мислитель Платон, називаючи так «мистецтво управління кораблем». Французький фізик і математик А.М. Ампер, ще на початку XIX століття, створюючи класифікацію наук, називає кібернетику наукою про управління державою. А ось в 1948 р. американський математик Норберт Віннер видає книгу «Кібернетика», в якій визначає це поняття «як наука управління і зв'язку у тварині і машині» (суспільство вже випадає з цього визначення). Теоретичну основу кібернетики, особливо технічної, становить теорія інформації, яка є молодшою за кібернетику [10, с. 6].

Основні частини теорії інформації розглядалися американським спеціалістом Р. Хартлі (1928 р.), а саме: в праці, присвяченій вимірюванню кількості інформації. Розвиток ідей вченого поклав початок теорії інформації, який вираховується із появи у 1948 р. фундаментальної роботи американського дослідника Клода Шеннона «Математична теорія зв'язку». В певному розумінні, до проблем теорії інформації відносять всі проблеми та задачі, до формулювання яких входить поняття інформації.

Слід відмітити видатних вчених, які зробили важливий внесок у розвитку окремих розділів теорії інформації: Н. Вінер,

Д. Міддлтон, А. Колмогоров, О. Хінчин, В. Фано, У. Пітерсон, Р. Боуз, Л. Бородін та ін.

Знання теоретичних основ інформатики є невід'ємною частиною теоретичної підготовки фахівців у галузі інформаційних технологій, тощо.

Посібник розроблено таким чином, щоб студенти могли послідовно вивчити всі основні поняття з теорії: інформацію, кодування, розпізнавання образів, автоматів, кібернетики, які викладені в даному посібнику, що допоможе студентам краще засвоїти навчальний матеріал.

Навчальний посібник призначений для вивчення дисципліни «Теоретичні основи інформатики», яка належить до групи нормативних і займає важливе місце у підготовці спеціалістів, магістрів за напрямом 111 Математика, 014.08 Середня освіта (Фізика), 014.04 Середня освіта (Математика), 014.09 Середня освіта (Інформатика).

Навчальне видання відповідає лекційному курсу дисципліни, який викладається автором в Національного університету «Чернігівський колегіум» імені Т.Г. Шевченка. Навчальне видання може бути використаним для студентів денної і заочно-дистанційної форм навчання. З цією метою, під час викладення теоретичного матеріалу наводяться питання для самоконтролю, завдання самостійної роботи, тестові завдання, питання до екзамену.

ЗМІСТОВИЙ МОДУЛЬ І. ТЕОРЕТИЧНІ ОСНОВИ ІНФОРМАТИКИ

Тема 1. Теорія інформації

План

1. Основи теорії інформації
2. Методологічні засади теорії інформації
3. Концепції теорії інформації
4. Ентропія

1. Основи теорії інформації

Теорія інформації – це розділ кібернетики, в якому за допомогою математичних методів вивчаються способи вимірювання кількості інформації, що міститься в будь-яких повідомленнях, способи кодування для економного подання повідомлень і надійної передачі їх каналами зв'язку із перешкодами [7, с. 19].

Теорія інформації об'єднує такі теоретичні напрями, як кількісна оцінка інформації, кодування повідомлень, і їх стиснення, оцінка ефективності та завадостійкості передачі кодованих повідомлень [7, с. 19].

Початком виникнення теорії інформації вважається 1948 рік, а її засновником – К. Шеннон (США), однак деякі результати, що увійшли до цієї науки, були отримані раніше Р. Хартлі (Англія, 1928 р.), В.А. Котельніковим (СРСР, 1933 р.), А.М. Колмогоровим (СРСР, 1941р.).

Важливою задачею теорії інформації є побудова оптимальних інформаційних систем, а об'єктом дослідження – всі складові таких систем. Процеси у інформаційних системах уявляються випадковими, а відповідним математичним апаратом є теорія ймовірності.

Основними поняттями теорії інформації є інформація, повідомлення, сигнали, знаки, коди, перешкоди (шуми), інформаційні системи, тощо.

Термін «*інформація*» походить від латинського «*informatio*», яке має три значення:

- роз'яснення, виклад, витлумачення;
- представлення, поняття;
- ознайомлення, просвіта [10].

Єдиного визначення інформації не існує. В залежності від галузі використання, термін «інформація» одержав безліч визначень.

Інформація (від лат *informo* – формувати, утворювати) у широкому розумінні – сукупність відомостей вербального і невербального характеру; у вузькому – сукупність повідомлень щодо актуальних питань, що стосуються будь-якої сфери діяльності людини або суспільства. У «Термінологічному словнику з інформатики», виданому Міжнародним центром з наукової та технічної інформації, інформація трактується як зміст зв'язку між матеріальними об'єктами, що проявляється у зміні станів цих об'єктів.

Законом України «Про інформацію» визначено, що «інформація – це документовані або публічно оголошені відомості про події та явища, що відбуваються в суспільстві, державі та навколишньому природному середовищі» [9].

Отже, поняття «інформація» багатозначне, серед найпоширеніших тлумачень можна назвати наступні:

- а) відомості або повідомлення про щось (побутове);
- б) оригінальність, новизна;
- в) комунікація та зв'язок, в процесі якого усувається невизначеність (теорія зв'язку, американський інженер *Клод Шеннон*); «Інформація – це упорядкування». Базові поняття цього визначення упорядкування – безлад – комунікація, відповідають засадам шеннонівської теорії інформація – ентропія – зв'язок. Під інформацією К. Шеннон розумів сигнали, корисні отримувачу. Некорисні сигнали – це шум, перешкоди. Інформація тут розглядається як те, що усуває невизначеність, що вимірюється кількістю невизначеності, яку вона усуває.

Шум – це випадкові збудження фізичної природи, які реєструються технічними системами [25].

Таким чином, під інформацією стали розуміти не всі повідомлення, а лише ті, що були новими і приносили користь для прийняття певних рішень.

- г) передача різноманітності (англійський філософ *Уільям Росс Ешбі*). Чим більше у певному об'єкті відмінних один від одного елементів, тим більше він містить інформації. Інформація є там, де є відмінність, принаймні, між двома об'єктами. Інформації немає, якщо елементи нерозрізнені. Її сутність – теорія інформації досліджує процеси «передачі різноманіття» каналами зв'язку,

причому «інформація не може передаватися у більшій кількості, ніж це дозволяє кількість різноманіття». На основі робіт К. Шеннона і Н. Вінера, У. Ешбі відкрив закон «необхідного різноманіття». Його сутність: для управління кібернетичною системою потрібен регулятор, який обмежує різноманіття збурень, що можуть зруйнувати систему. Однак регулятор допускає таке їх різноманіття, яке корисне для системи.

д) універсальна субстанція, що пронизує всі сфери людської діяльності, слугує провідником знань та думок, інструментом спілкування, взаєморозуміння та співробітництва, утвердження стереотипів мислення та поведінки (ЮНЕСКО);

е) документовані або публічно оголошені відомості про події та явища, що відбуваються у суспільстві, державі та навколишньому природному середовищі (Закон України «Про інформацію»).

Під *даними* розуміють усі відомості, здобуті від навколишнього світу та подані у формалізованому вигляді (літерами, цифрами, символами тощо).

Дані – отримані відомості, подані у формалізованому вигляді (літерами, цифрами, символами тощо) [26 Дмитриев]. Дані обробляються комп'ютерами, для передачі даних використовуються комп'ютерні мережі, дані зберігаються в базах даних. Дані – сировина для створення інформації: перетворення й обробка даних дозволяє (з точки зору теорії інформації) витягнути новизну та перетворити її на знання.

Дані, що підлягають передачі, називаються повідомленнями. Повідомлення стають інформацією тільки в момент їх застосування, тобто інформація – це використовувані повідомлення, причому такі, які відзначаються новизною і раніше не були відомі одержувачу. Є й інше, більш широке визначення інформації, згідно з яким *інформація* – це відомості, що є об'єктом зберігання, передавання та перетворення [8, с. 8].

Семіотика, або семіологія (від грец. semeion – знак, ознака) – наука, яка досліджує способи передачі інформації, властивості знаків та знакових систем в людському суспільстві (головним чином природні та штучні мови, а також деякі явища культури, системи міфів, ритуалів), природі (комунікація у тваринному світі) або в самій людині (зорове та слухове сприйняття тощо) (Рис.1.1).

Для можливості технічного оброблення (передачі, запису та ін.), повідомлення має бути перетворене на *сигнал* – матеріальний носій, що відображає повідомлення.

Відрізняють сигнали:

- звукові (акустичні)
- електричні
- оптичні
- гідравлічні

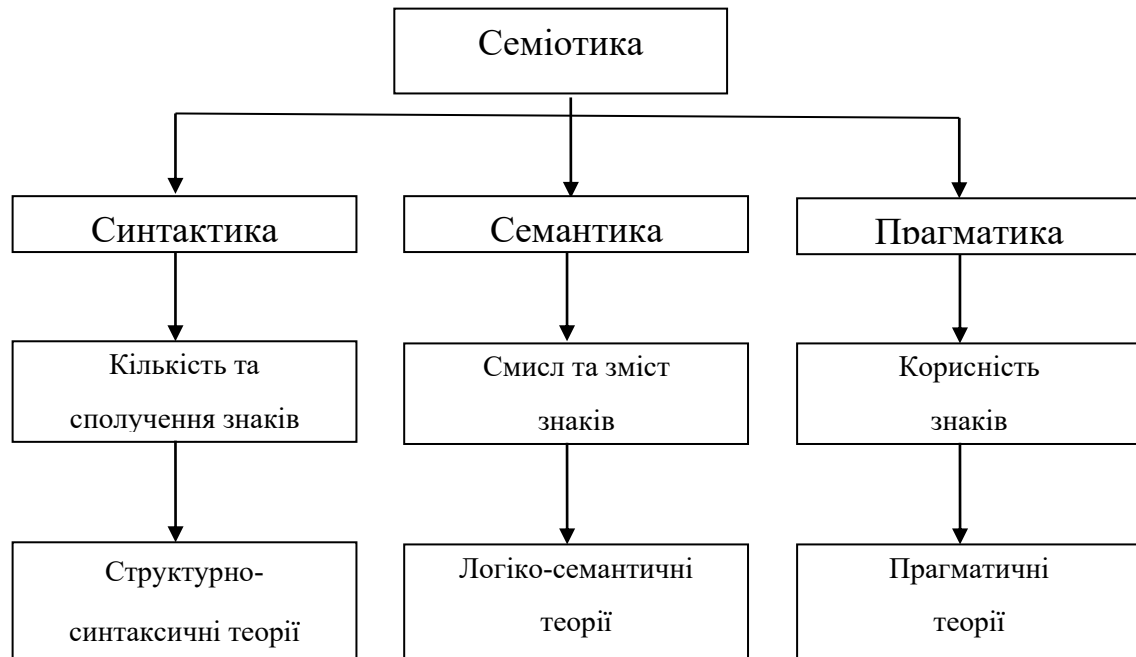


Рис. 1.1. Концепція інформації в рамках семіотичного підходу

Один вид сигналу можна перетворювати на інший (електричний на звуковий, оптичний на електричний тощо).

Будь-який сигнал характеризується такими основними параметрами:

- тривалістю;
- шириною частотного спектра;
- динамічним діапазоном [8, с. 8].

Під *тривалістю* T_c сигналу розуміють час, протягом якого він знаходиться в каналі зв'язку. Частотний спектр F_c сигналу визначає смугу частот, яку він охоплював під час передачі каналом зв'язку. Залежно від виду сигналу (аналоговий, дискретний), *частотний спектр* може бути і нескінченним; тому на практиці його обмежують для можливості передачі каналами з обмеженою смугою частот.

Середньою потужністю P_c сигналу є потужність, яка забезпечується апаратурою під час його знаходження до каналу зв'язку. Частіше, замість P_c користуються поняттям *динамічного діапазону* D_c , що визначається логарифмом відношення найбільшої (максимальної) миттєвої потужності сигналу ($P_{c\max} = P_c$) до найменшої (мінімальної) $P_{c\min}$, дозволене значення якої дорівнює потужності завад:

$$(P_{c\min} = P_3):$$

$$D_c = \log(P_c / P_3)$$

Усі ці параметри сигналу є його обсягом:

$$V_c = T_c F_c D_c.$$

Повідомлення та відповідні сигнали можуть бути неперервними (аналоговими) та дискретними (знаковими). Перші описуються неперервною функцією часу [8, с. 9].

Дискретне повідомлення – скінченна послідовність окремих символів (знаків, літер) з обмеженою тривалістю. Для перетворення дискретного повідомлення на сигнал потрібна операція кодування.

Неперервні повідомлення можна передавати дискретними методами: неперервні сигнали, якими передаються ці повідомлення, перетворюються на дискретні за допомогою операцій квантування за рівнем та дискретизації в часі. На приймальному боці виконується обернене перетворення: за прийнятими дискретними сигналами відновлюються передані неперервні сигнали.

Дискретні сигнали, як засіб передачі повідомлень, більш поширені, ніж неперервні, завдяки тому, що вони меншою мірою зазнають впливу завад і спотворень в каналах зв'язку, а в разі спотворення – їх легше відновити.

Каналом зв'язку називається сукупність технічних засобів, що забезпечує передачу повідомлень від джерела до одержувача [8, с. 10].

Види інформації

За критерієм форми інформація поділяється на два види [11, с.26]:

а) дискретна форма – це послідовність символів, що характеризує дискретну, змінну величину (кількість дорожньо-транспортних пригод);

б) аналогова форма – це величина, яка характеризує процес, що не має перерв або інтервалів (температура тіла людини).

За сферою виникнення інформацію виокремлюють як:

а) елементарну (механічну), що відображає процеси, явища неживої природи;

б) біологічну, що відображає процеси тваринного і рослинного світу;

в) соціальну, що відображає процеси людського суспільства.

За засобом передачі і сприйняття:

а) візуальна;

б) аудіальна;

в) тактильна;

г) органолептична (смаки і запахи);

д) технічна (видається і сприймається технічними засобами).

Інформацію, що створюється і використовується людиною, за суспільним позначенням можна поділяти на *три види*:

а) особиста – зорієнтована на конкретну людину;

б) масова – зорієнтована на будь-яку людину, що бажає долучитися і скористатися інформацією;

в) спеціальна – призначена для використання вузьким колом осіб, що займаються вирішенням складних спеціальних завдань в галузі науки, техніки, економіки.

За засобами кодування виокремлюють наступні типи інформації:

а) символічну, ґрунтовану на використанні символів (літер, цифр, знаків, тощо);

б) текстову, ґрунтовану на використанні комбінацій символів (книги, журнали, тощо);

в) графічну, ґрунтовану на використанні вільного поєднання у просторі графічних примітивів. До цієї форми залучаються фотографії, схеми, креслення.

З точки зору системного підходу виокремлюють:

а) внутрішню інформацію (структурну) – притаманна всім об'єктам живої і неживої природи і виникає внаслідок відбору, фіксації і закріплення структурних видозмін у системі позитивного досвіду взаємодії із зовнішнім середовищем. Як приклад, квітка у темній кімнаті (цикл розкривання і закривання пелюсток регулюється внутрішньою структурою).

б) зовнішню інформацію – інформацію як засіб організації будь-якої системи, тісно пов'язаної із відображенням впливу

одного об'єкту на інший. Приклад впливу зовнішньої інформації на біологічні системи – ритми сонячної активності, місячні ритми, тощо.

За місцем виникнення:

- а) вхідна – це інформація, яка поступає до об'єкту;
- б) вихідна – інформація, яка поступає від об'єкту до іншого.

За стадіями обробки інформації:

- а) первинна – це інформація, яка виникає безпосередньо в процесі діяльності на початковій стадії;
- б) вторинна – це інформація, яка виникає внаслідок обробки первинної інформації і може бути проміжною та результативною;
- в) проміжна – використовується як вихідні дані для наступних дій;
- г) результативна – виникає в процесі обробки первинної, проміжної інформації та використовується для прийняття рішень.

2. Методологічні засади теорії інформації

На сучасному етапі *теорія інформації* містить такі *розділи* як:

- аксіоматична база;
- принципи вимірювання кількості інформації;
- кодування інформації;
- передача інформації.

Кібернетична система містить:

- пристрої для збору інформації;
- пристрої для кодування/декодування інформації;
- пристрої для передачі інформації;
- пам'ять для зберігання інформації.

Виділяють основні види керуючих систем:

- механічні;
- електричні;
- електронні (комп'ютерні);
- біологічні (в тому числі людина);
- соціальні.

Властивості (показники якості), що характеризують її основні споживацькі якості:

1. об'єктивність – незалежність інформації від методів її отримання, фіксації, чиєїсь думки чи судження;
2. повнота – достатність інформації для прийняття

правильного рішення;

3. достовірність (точність) – ступінь відповідності інформації, що характеризує певний об'єкт його реальному стану;

4. актуальність (своєчасність) – ступінь відповідності деякої інформації поточному моменту часу;

5. цінність (корисність) – міра значущості певної інформації для прийняття певного рішення;

6. доступність (розпізнаність) – міра можливості отримання тієї чи іншої інформації та її сприйняття;

7. захищеність – неможливість несанкціонованого використання чи зміни інформації;

8. ергономічність – зручність форми або об'єму інформації;

9. фасцинація – привабливість інформації.

2.1. Кількісна оцінка інформації

У інформатиці для виміру даних використовується той факт, що різні типи даних мають універсальне двійкове представлення. Мінімальною одиницею інформації у комп'ютері є **біт**. Біт може бути або виключеним, при цьому його значення дорівнює 0, або включеним, тоді його значення дорівнює 1. Група з 9 бітів уявляє собою байт, 8 біт якого містять дані, і 1 біт – контроль на парність. 8 біт дають 256 різних комбінацій включених та виключених станів від 0000000 до 11111111.

Вимога контролю на парність полягає у тому, що кількість включених бітів у байті завжди повинна бути непарною. Біти у байті пронумеровані від 0 до 7 справа наліво. Число 210 дорівнює 1024, що становить 1 Кбайт та позначається буквою К.

Процесор у комп'ютері використовує 16-бітову архітектуру. 16-бітове (двобайтове) поле називається *словом*. Біти у слові пронумеровані від 0 до 15 справа наліво.

Для кількісної оцінки інформації вдаються до теорії ймовірностей, що в свою чергу, базується на теорії множин. Не заглиблюючись далі, акцентуємо: щоб визначити множину, достатньо вказати характеристичну властивість елементів, тобто властивість, що притаманна всім елементам цієї множини і лише їм. (Приклад: множина точок [елементів] даної лінії [множини]). Іншими словами: множина відноситься до категорії якості, а не кількості. Місток між якістю (ідеальним) та кількістю (матеріальним) прокладають наступні аксіоми:

1. Кожній випадковій події А поставлене у відповідність

невід'ємне число $P(A)$, що називається його ймовірністю. (чомусь ідеальному можна поставити у відповідність щось з природи матеріального).

2. Ймовірність всієї множини $P(E) = 1$.

3. Якщо події $A_1, A_2, A_3, \dots, A_n$, попарно несумісні, то $P(A_1 + A_2 + A_3 + \dots + A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n)$. Цього досить для розуміння подальших міркувань з приводу вимірювання інформації. В якості основної характеристики повідомлення в теорії інформації прийнято величину, що називається *кількість інформації*. Це поняття пов'язане зі *ступенем невизначеності*. Нехай алфавіт джерела повідомлень становить m знаків (елементів), кожний з яких може бути елементом повідомлення. Кількість N можливих повідомлень довжини n дорівнює числу перестановок з необмеженими повтореннями: $N = m^n$.

Якщо для одержувача всі N повідомлень від джерела є *рівноймовірними*, то одержання конкретного повідомлення рівносильне для нього випадковому вибору одного з N_i повідомлень з ймовірністю $P_i = 1/N$. Тоді кількість інформації можна виразити через ймовірності надходження повідомлень (про який ми будемо говорити пізніше).

Одиницю кількості інформації на один елемент повідомлення називають *двійковою одиницею або БІТОМ*. Тобто одиниця невизначеності (двійкова одиниця або біт) є невизначеністю вибору з двох рівноймовірних подій (bit – скорочення від англ. binary digit – двійкова одиниця).

Оскільки з того, що $\log_2 m = 1$, випливає що $m = 2$, зрозуміло: *1 біт – це кількість інформації, котрою характеризується один двійковий елемент при рівноймовірних станах 0 і 1 (НІ або ТАК)*. Математичною моделлю біта є однорозрядне двійкове число, ентропія та інформація якого дорівнюють 1-му біту.

Зі шкільного курсу інформатики Вам можуть згадатися інші визначення:

1 біт – кількість інформації, що міститься в повідомленні, яка зменшує невизначеність знань в 2 рази;

1 біт – кількість інформації, яка дозволяє вибрати правильний варіант з двох можливих.

У обчислювальній техніці бітом називають найменшу «порцію» пам'яті, необхідну для збереження одного з двох символів «0» і «1», що використовуються для внутрішнього представлення даних і команд.

Двійкове повідомлення довжиною n містить n біт інформації. Очевидно, що для практичного використання необхідна «зручніша» одиниця виміру інформації, тому для виміру комп'ютерної інформації застосовується *восьми бітове число* – *Байт* – це одиниця кількості інформації, що дорівнює 8 бітам.

З того ж таки шкільного курсу інформатики: *байт* – група з 8 біт. Одним байтом можна виразити 256 різних значень (від 0 до 255).

Байт – вісім бінарних розрядів (бітів) – приймає 2^8 значень: від 0 до 255.

Тетрада (півбайт, ніббл) – чотири двійкових розряди, половина байта – приймає 2^4 значень: від 0 до 15.

Слово – код, що складається з кількох байтів (найчастіше всього 2 байти – 16 розрядів, 4 байти – 32 розряди, 8 байт – 64 розряди).

Сучасні інформаційні засоби оперують з величезними об'ємами інформації, тому не зайве «мати перед очима» таблицю префіксів. Якщо за основу логарифма прийняти 10, то ентропія виразиться в десяткових одиницях на один елемент повідомлення – *дітах*, тобто $1 \text{ діт} = \log_{10} 2 \text{ біт} = 3,32 \text{ біт}$.

Передача інформації

Канал зв'язку передає інформацію в просторі (не в часі) (Рис. 1.2).

Кожен канал передачі інформації характеризується певною пропускнуою здатністю, тобто кількістю інформації, яку він може передати за одиницю інформації при умові, що приймач інформації зможе її правильно розкодувати. Вимірюють у кількості бітів за секунду (біт/с) — бод.

Види каналів зв'язку:

- а) для людини;
- б) для технічних пристроїв.

Канали зв'язку для людини: повітря, світлові промені (80%), тощо.

Крім того, людина може використовувати й технічні канали зв'язку.

Способи *передачі інформації*:

- послідовна передача цілого повідомлення, передача порціями (пакетами).
- Постійна передача: радіо, телебачення.

- Передача порціями (пакетами): комп'ютерна техніка.



Рис. 1.2. Загальна схема передачі інформації

При передачі інформації каналами зв'язку існує проблема її спотворення, викривлення внаслідок різноманітних чинників, а тому, відповідно зростає необхідність дотримання найвищої степені достовірності та відповідності інформації що передається.

Достовірність інформації вимірюють у кількості помилок на фіксовану порцію інформації, наприклад на 1000 знаків. Для підвищення достовірності разом із кожною порцією інформації передають також контрольні розряди.

Приклад перевірки за допомогою контрольного розряду: всі коди порції інформації сумують, результат ділять на 11, а частку від ділення записують у порції як останнє число (це – контрольний розряд) й передають приймачеві. Розкодовуючи, приймач перевіряє, чи контрольний розряд (частка від ділення), який отримав він, дорівнює тому, що надіслав передавач. Якщо він не рівний, то це означає, що під час передачі, внаслідок шумів, десь виникло спотворення повідомлення, а тому цю порцію інформації потрібно передати ще раз (передача повторюється доти, доки не буде отримано збігу контрольних розрядів).

Максимальна довжина контрольних розрядів може дорівнювати повторній передачі повідомлення (контроль методом дублювання).

3. Концепції теорії інформації

3.1. Імовірнісна (класична) концепція теорії інформації

Класична теорія інформації Р. Хартлі та К. Шеннона базується на теорії імовірностей (ТІ). Кількість інформації для рівноймовірних здійснень випробувань в цій теорії вимірюється як логарифм від кількості можливих варіантів (N) завершення випробування:

$$I = \log_2 N = \log_2 m^n = n \log_2 m \quad (1.1)$$

де N – кількість можливих варіантів здійснень випробування, тобто кількість можливих повідомлень; n — кількість розрядів у повідомленні (наприклад, 3); m — кількість символів в алфавіті (наприклад, два: 0 та 1).

Одиницею вимірювання кількості інформації служить біт.

Фундаментальна формула була вдосконалена К. Шенноном для випадку, коли здійснення подій не є рівноймовірним: оскільки формула (1.1) Р. Хартлі $I = \log N$ внаслідок звичайних математичних перетворень дорівнює $I = -\log(1/N)$ де $(1/N)$ є ймовірністю (p) здійснення випробування, то можна записати, що $I = -\log p$. Якщо ж здійснення подій нерівноймовірне, то для варіанту завершення і кількість інформації в ньому становитиме:

$$I_i = -p_i \log_2 p_i. \quad (1.2)$$

Використання даної формули показує, що чим менша ймовірність здійснення варіанту і певного випробування, тим більше інформації він несе.

Кількість інформації у всіх випробуваннях експерименту визначають як арифметичну суму по всіх здійсненнях i .

Дана математична модель досконало працює для тих випадків, коли мова йде про кодування інформації, тобто при передачі інформації каналами зв'язку чи її зберіганні на носіях інформації. Але вона зовсім не працює, коли йдеться про визначення кількості інформації в об'єктах, позначених цими кодами.

Приклад. Визначаючи кількість інформації в літері Р, ця теорія зовсім не аналізує, що ця літера, по-перше, складається з двох компонентів: прямої лінії та півкола; по-друге, пряма та півколо з'єднані між собою; по-третє, товщина лінії, яка утворює півколо,

в різних точках є різною; по-четверте, пряма лінія нахилена під певним кутом, тощо.

Приклад. Вимірюючи кількість інформації у слові *мама*, ця теорія визначає лише кількість інформації в літерах цього слова, тобто $I_{\text{мама}} = (-pm * \log pm) + (-pa * \log pa) + (-pa * \log pa) + (-pm * \log pm)$, проте зовсім не визначає, скільки інформації міститься в значенні цього слова (*мама* — людина, *мама* — жінка, *мама* — має двоє очей, ніг, рук і т. д., а конкретна моя *мама* має ще й певний вираз обличчя, колір очей, ріст, вагу тощо). Зрозуміло, що семантична інформація (що таке *мама*), тут повністю проігнорована, так само, як інші види інформації, що є в текстах — синтаксична, стилістична, граматична тощо.

Приклад. Візьмемо дві сторінки газети. На першій — всі літери записані в ланцюжки на основі даних генератора випадкових чисел (з урахуванням частот появи літер і впливу на них контексту як це має місце в реальних текстах); в ілюстраціях, пікселі (з чорним чи білим зображенням) записані на площині в порядку, заданому на основі того самого генератора випадкових чисел. При цьому, вказана сторінка нічого спеціально не шифрує, тобто ключа до її розшифрування, в принципі, немає. На другій сторінці газети — осмислені тексти; ілюстрації є фотографіями об'єктів реального світу. Коли порівняти кількість інформації на цих двох сторінках за Шеннонівською теорією інформації, то виявиться, що вони є практично однаковими, хоча кожна людина з власного досвіду знає, що це не так.

3.2. Алгоритмічна концепція теорії інформації

Друга концепція була опублікована в 1965 р. відомим російським математиком А. Колмогоровим. Вона отримала назву алгоритмічної, оскільки передбачала визначення кількості інформації на основі програми (і, відповідно, алгоритму), яка перетворює один об'єкт в інший. Кількість інформації за теорією інформації, що базується на алгоритмах перетворення (далі — алгоритмічна ТІ), вимірюється як довжина програми, що дає змогу перетворити об'єкт А в об'єкт В, тобто:

$$I = L[G(A, B)], \quad (1.3)$$

де L — функція отримання довжини програми (в байтах чи бітах), а G — програма перетворення.

У поданому вище прикладі алгоритмічна ТІ дасть цілком інший результат. Згідно з її методикою, кількість інформації в літері Φ стосовно P буде рівною, наприклад, кількості байтів програми, яка перетворює літеру Φ в P . Тут, правда, можливі різні результати залежно від того, що вибрати за елемент перетворення: коли цілі літери, то кількість інформації буде найменшою; коли стандартизовані елементи літер — середньою; коли пікселі — максимальною. Виміряти ж кількість інформації в будь-якій літері, взятій окремо, взагалі неможливо, хіба що порівнювати її з пробілом. Розглядати ж можливість перетворення літери на себе саму немає сенсу, оскільки тоді довжина програми перетворення рівна нулю, а отже, й кількість інформації в ній виявиться також рівною нулю.

Спосіб вимірювання кількості інформації в значенні слова *мама* за цією теорією — невизначений (на нашу думку, неможливий).

Вадами цієї теорії є те, що вона:

а) в методі вимірювання, як параметр, використовує довжину алгоритму, який залежить від кількості «будівельних елементів», що дають змогу перетворити один об'єкт на інший (чим дрібніші «будівельні елементи», тим довшою виявляється програма перетворення для однієї й тієї ж пари об'єктів);

б) не має методу, який дає змогу виміряти кількість інформації в одному об'єкті, оскільки для їх вимірювання завжди потрібно як мінімум два;

в) не враховує, що одна й та ж програма може бути призначена для перетворення не одного, а цілого класу об'єктів; при цьому така програма матиме сталу довжину, хоча для різних об'єктів кількість виконаних команд може бути різною.

Образна концепція теорії інформації як різновид алгоритмічної концепції теорії інформації.

У будь-якому повідомленні M кількість інформації в образі O (I_o) будемо вимірювати за напівемпіричною формулою, яка узгоджується з положеннями інформаційних підходів А. Колмогорова:

$$I_o = q^{-1} L[G(O_1, O_2)], \quad (1.4)$$

де q — ймовірність правильного розпізнавання образу O_1 кібернетичною системою;

L — функція отримання довжини розгорнутої програми;

G — програма розпізнавання образу, біти;

O_2 — еталонний образ кібернетичної системи.

Тут імовірність q визначається апостеріорно, за допомогою багаторазових експериментів.

Для простішого оперування у формулі (1.4) величину $L[G(O_1, O_2)]$, позначимо як Q , тобто прийнемо, що $L[G(O_1, O_2)] = Q$. Тоді формулу (3.4) можна скорочено записати як:

$$I_o = q^{-1} Q. \quad (1.5)$$

Саме в такому вигляді й будемо використовувати дану формулу.

Інтерпретація інформації як розпізнаного образу.

Нерозпізнані образи не є інформацією. Такі образи — це лише відображення.

3.3. Семантична концепція теорії інформації

Семантичний аспект передає смисловий зміст інформації і співвідносить її з раніше одержаною інформацією. Рудольф Карнап та Йошуа Бар-Хілел поклали початок застосування ідеї та методів імовірнісної логіки та логічної семантики для аналізу інформаційного змісту мови науки.

Вимірювання кількості інформації в семантичній теорії інформації, що базується на математичній логіці (логічна ТІ), можливе лише для тверджень за формулою:

$$I = \log_2 F(s), \quad (1.6)$$

де F — функція отримання кількості інформації, що залежить від кількості можливих станів у складових твердження; S — саме твердження.

В цій теорії виміряти кількість інформації в літері P або в значенні окремого слова *мама* взагалі неможливо, оскільки літера чи слово подані окремо, поза складом тверджень.

Вадами цієї теорії є те, що вона:

а) не дає можливості виміряти кількість інформації як у словах (семантичну інформацію), так і в тексті (тобто інформацію, що є в тексті за межами тверджень);

б) не дає можливості виміряти кількість невербальної інформації, що свідчить про її незаперечну обмеженість.

Близькою за своєю суттю до семантичної теорії інформації є концепція М. Мазура.

4. Ентропія

Поняття ентропії в теорії інформації вперше запропонував *К. Шеннон*. Завдяки вченому, ентропія стала використовуватися як міра корисної інформації у процесах передачі сигналів. Слід нагадати, що під інформацією *К. Шеннон* розумів сигнали необхідні користувачеві. Некорисні сигнали – це шум, звукові перешкоди.

Шум - це завади в каналі зв'язку.

Ентропія – кількість інформації, яка приходить на один символ [7, с. 30].

Якщо сигнал на виході з каналу зв'язку є точною копією сигналу на вході, це означає відсутність ентропії. Відсутність шуму означає максимум інформації.

Термін *негентронія* (негативна ентропія) з'явилося у 30-х рр. минулого століття. *Е. Шредингер* дійшов висновку, що біологічним системам для свого існування слід здобувати з оточуючого середовища негативну ентропію для компенсації внутрішнього виробництва ентропії та загальмувати рух до сфери термодинамічної рівноваги, тобто смерті. Порівнюючи інформацію та ентропію, *Л. Бріллюєн* сформулював негентропійний принцип інформації: «інформація є негативним вкладом в ентропію». Бельгійський вчений *І. Пригожин* отримав нобелівську премію за встановлення стійких структур (у хімії, біології) з негативною ентропією.

А. Моль, продовжуючи дослідження *К. Шеннона* у сфері інформації, підкреслював: інформація та її кількість є виміром складності структур, ступенем непередбачуваності, змістовної новизни. Банальне повідомлення завжди передбачуване. Виміром інформаційної насиченості є ступінь його непередбачуваності [12, с. 19].

Отже, *шум* – це все, що маскує корисний сигнал. Шум призводить до неповноти, або спотворення отриманої інформації, знижуючи її якість.

Кількість інформації, яка припадає на одне повідомлення, характеризує дискретне джерело повідомлень в цілому. Інше

джерело з іншим ансамблем повідомлень матиме зовсім іншу питому кількість інформації. Ця загальна характеристика джерела повідомлень називається його ентропією $H(A)$. Вона має фізичний зміст середньостатистичної міри невизначеності відомостей спостерігача A відносно стану спостережуваного об'єкта.

Точно ентропію можна визначити як математичне сподівання питомої кількості інформації:

$$H(A) = \sum_i p(a_i) I(a_i) = -\sum_i p(a_i) \log p(a_i).$$

Таблиця 1.1

$a_i \in A$	a_1	a_2	a_3	a_4	a_5	a_6
$p_i \in P$	0.4	0.3	0.15	0.1	0.03	0.02
$I(a_i), \text{біт}$	1.322	1.737	2.737	3.322	5.059	5.644

Згідно з даними табл. 1.1 маємо
 $H(A) = 0,4 \cdot 1,332 + 0,3 \cdot 1,737 + 0,15 \cdot 2,737 + 0,1 \cdot 3,322 + 0,03 \cdot 5,059 + 0,02 \cdot 5,644 = 2,0573$
біт/повідомлення.

Необхідні витрати енергії на передачу повідомлення пропорційні його ентропії (середній кількості інформації на одне повідомлення). Виходить, що кількість інформації в послідовностях визначається кількістю повідомлень N у послідовності та ентропією $H(A)$ джерела, тобто:

$$i(N) = NH(A).$$

Властивості ентропії:

1) ентропія дорівнює нулю, якщо хоча б одне з повідомлень достовірне;

2) ентропія завжди більша або дорівнює нулю, є величиною дійсною і обмеженою;

3) ентропія джерела з двома альтернативними подіями може змінюватися від 0 до 1;

4) ентропія – величина адитивна: ентропія джерела, повідомлення якого складаються з повідомлень декількох статистично незалежних джерел, дорівнює сумі ентропії цих джерел;

5) ентропія максимальна, якщо всі повідомлення мають однакову імовірність. Таким чином, якщо алфавіт $A = \{a_1, a_2, \dots, a_k\}$ має k різних повідомлень, то $H(A) \leq \log k$. Тут рівність стосується

тільки рівномірних і статично незалежних повідомлень $a_i \in A$. Число k називається обсягом алфавіту повідомлень.

Безумовна ентропія

Термін «безумовна ентропія» запозичений з математичної статистики за аналогією з безумовною ймовірністю, що стосується статистично незалежних подій, тут – повідомлень.

Безумовна ентропія це питома кількість інформації на один елемент повідомлень, складених з алфавітів, між символами яких немає взаємозалежності.

Коли говорять про безумовну ентропію, то слово «безумовна» опускають. Оговорюють інші види ентропії як менш поширені.

Ентропія джерела повідомлень $H(A) = -\sum_{i=1}^m p(a_i) \log p(a_i)$ і

являє собою невизначеність появи на виході джерела повідомлень букви первинного алфавіту.

Ентропія приймача повідомлень $H(B) = -\sum_{i=1}^m p(b_i) \log p(b_i)$ являє собою невизначеність появи на вході приймача букви після її появи на виході джерела повідомлень.

Таким чином, основними **властивостями безумовної ентропії** дискретних повідомлень є такі:

1. Якщо відомо, що подія точно відбудеться (це еквівалентно передачі повідомлення з однією якісною ознакою), то його ентропія рівна нулю.
2. Якщо відомо, що подія точно не відбудеться (це еквівалентно відсутності i -ої якісної ознаки), то його ентропія рівна нулю.
3. Якщо дослід являє собою послідовність рівноімовірних подій (це еквівалентно передачі повідомлень з рівноімовірними взаємонезалежними символами), то ентропія є максимальною.
4. Якщо подія з рівною імовірністю може відбутися або не відбутися, то його ентропія максимальна.
5. Ентропія зі скінченною кількістю результатів завжди є додатною величиною.
6. Ентропія бінарного відношення володіє властивістю симетрії: збільшення або зменшення ймовірності будь-якого з елементів бінарного повідомлення на одну і ту ж величину приводить до однакової зміни ентропії.

7. Ентропія складного повідомлення, яке складається з деяких часткових повідомлень, рівна сумі ентропій складових повідомлення: $H(AB) = H(A) + H(B)$.
8. Ентропія джерел, дорівнює сумі ентропії цих джерел [8, с.32].

Умовна ентропія

При передачі змістовних повідомлень одні букви зустрічаються частіше, інші рідше, одні букви і слова часто ідуть за іншими.

Наприклад, в англійській мові найчастіше використовується буква *e*; у французькій після букви *q* завжди іде буква *u*, якщо *q* не стоїть в кінці слова. Щодо взаємодіючих систем, то переважно стан однієї з них впливає на стан іншої. В таких випадках ентропія не може визначатись лише на основі безумовних ймовірностей.

При обчисленні середньої кількості інформації на символ повідомлення взаємозалежність враховують через умовні ймовірності появи одних подій відносно інших, а отриману при цьому ентропію називають умовною ентропією.

Умовна ентропія застосовується при визначенні взаємозалежності між символами первинного алфавіту, для визначення втрат при передачі інформації по каналах зв'язку, при обчисленні ентропії об'єднання.

Наприклад, розглянемо текстове повідомлення, якщо при передачі n повідомлень символ A з'явився m раз, символ B з'явився l раз, а символ A разом з символом B – k раз, то ймовірність появи символу A $p(A) = m/n$; ймовірність появи символу B $p(B) = l/n$; ймовірність сумісної появи символів A та B $p(AB) = k/n$.

Умовна ймовірність появи символу A відносно символу B $p(A/B) = \frac{p(AB)}{p(B)} = \frac{k}{l}$ і умовна ймовірність появи символу B відносно символу A $p(B/A) = \frac{p(AB)}{p(A)} = \frac{k}{m}$.

Якщо відомо умовну ймовірність, то можна легко визначити й ймовірність сумісної появи символів, використовуючи попередні формули:

$$p(AB) = p(B)p(A/B) = p(A)p(B/A).$$

Розглянемо формули умовної ентропії, використовуючи формулу Шеннона:

$$H(b_j/a_i) = -\sum_j p(b_j/a_i) \log_2 p(b_j/a_i) \quad (1.7)$$

$$H(a_i / b_j) = -\sum_i p(a_i / b_j) \log_2 p(a_i / b_j) \quad (1.8)$$

де індекс i вибрано для характеристики довільного стану джерела повідомлень A , індекс j вибрано для характеристики довільного стану адресату B .

Розрізняють поняття часткової та загальної умовної ентропії. Вирази (1.7) та (1.8) є **частковими умовними ентропіями**.

Загальна умовна ентропія повідомлення B відносно повідомлення A характеризує кількість інформації, яка міститься у будь-якому символі алфавіту, і визначається як сума ймовірностей появи символів алфавіту помножена на невизначеність, яка залишається після того як адресат прийняв сигнал

$$H(B / A) = -\sum_i p(a_i) H(b_j / a_i) = -\sum_i \sum_j p(a_i) p(b_j / a_i) \log_2 p(b_j / a_i) \quad (1.9)$$

Вираз (1.9) є загальним для визначення кількості інформації на один символ повідомлення для випадку нерівноімовірних та взаємозв'язаних символів.

Отже $p(a_i) p(b_j / a_i)$ являється імовірністю сумісної появи двох подій $p(a_i, b_j)$, то формулу (1.10) можна записати наступним чином:

$$H(B / A) = -\sum_i \sum_j p(a_i, b_j) \log_2 p(b_j / a_i) \quad (1.10)$$

Розглянемо **властивості умовної ентропії**:

1. Якщо ансамблі повідомлень A та B взаємонезалежні, то умовна ентропія A відносно B рівна безумовній ентропії A і навпаки:

$$H(A / B) = H(A); \quad H(B / A) = H(B).$$

2. Якщо ансамблі повідомлень A та B настільки жорстко статично пов'язані, що поява одного з них означає обов'язкову появу іншого, то їх умовні ентропії рівні нулю:

$$H(B / A) = H(A / B) = 0 \quad [8, \text{с. 34}].$$

Ентропія об'єднання

Ентропія об'єднання використовується для обчислення ентропії сумісної появи статистично залежних повідомлень. Наприклад, передаючи сто раз цифру 5 по каналу зв'язку із завадами, зауважили, що цифра 5 була прийнята 90 раз, цифра 6 – 8 раз і цифра 4 – 2 рази. Невизначеність виникнення комбінацій виду 5-4, 5-5, 5-6 при передачі цифри 5 може бути описана за допомогою ентропії об'єднання.

Ентропія об'єднання $H(A, B)$ – це невизначеність того, що буде посилатись A , а прийматись B . Для ансамблів переданих повідомлень та прийнятих повідомлень ентропія об'єднання являє собою суму вигляду

$$H(A, B) = -\sum_i \sum_j p(a_i, b_j) \log_2 p(a_i, b_j) \text{ біт / двасимволи} \quad (1.11)$$

Ентропія об'єднання та умовна ентропія пов'язані між собою наступними співвідношеннями:

$$H(A, B) = H(A) + H(B / A) = H(B) + H(A / B)$$

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

Властивості ентропії об'єднання:

1. Властивість симетрії: $H(A, B) = H(B, A)$.
2. При відсутності статистичної залежності між елементами ансамблів A та B умовні імовірності перетворюються на безумовні, тоді

$$H(A, B) = H(A) + H(B).$$

3. При повній статистичній залежності між елементами ансамблів A та B (наприклад, коли результат однієї події однозначно визначає інформацію про другу подію) умовні ентропії рівні нулю, тому $H(A, B) = H(A) = H(B)$ [8, с.45].

Питання для самоконтролю

1. Що таке повідомлення?
2. У чому полягає різниця між даними, повідомленнями та інформацією?
3. Властивості інформації.
4. Як здійснюється передача інформації?
5. Як визначається кількість інформації в одному повідомленні?
6. Що таке ентропія та які її властивості?
7. Назвіть концепції теорії інформації.

Тема 2. Теорія кодування

План

1. Характеристики кодів
2. Приклади бінарних кодів без надлишковості
3. Основні теореми кодування
4. Оптимальне кодування за відсутності шумів. Код Шеннона-Фано
5. Принципи завадостійкого кодування
6. Шифрування інформації

1. Характеристики кодів

Кодування інформації – процес присвоєння інформації умовних позначень за певним правилом; процес перетворення інформації з форми, зручної для безпосереднього використання інформації у форму, зручну для передачі, зберігання або автоматичної переробки (цифрове кодування, аналогове кодування, таблично-символьне кодування, числове кодування).

Завдання теорії кодування – здобуття ефективних алгоритмів кодування для джерел повідомлень і передачі даних по каналах зв'язку.

Код – сукупність знаків (символів) і система певних правил, за допомогою якої інформація може бути представлена (закодована) у вигляді набору з таких символів для передавання, обробки й зберігання.

Скінчена послідовність кодових знаків називається словом.

Декодування – операція, обернена до кодування, тобто відновлення інформації із закодованого вигляду (відновлення в первинному алфавіті в отриманні послідовних кодів).

Шифрування – різновид кодування.

Шифрування – це кодування даних з метою захисту від несанкціонованого доступу.

Шифр – код, значення та правила використання якого відомо обмеженому колу осіб.

Ключ – це конкретний секретний стан деяких параметрів алгоритму криптографічного перетворення даних, що забезпечує вибір тільки одного варіанту з усіх можливих для даного алгоритму. Зазвичай ключ являє собою послідовний ряд символів алфавіту.

Люди кодували повідомлення з давніх давен. Наприклад, з історії добре відомий шифр Цезаря, Азбука Морзе, шрифт Брайля, семафорна азбука (її використовують моряки). Сучасні людина часто зустрічається із закодованими повідомленнями у повсякденному житті. Наприклад:



Для того, щоб зрозуміти початкове повідомлення маючи закодоване, ми повинні знати правила (алгоритми) кодування. Вони можуть бути дуже різними. Якщо алгоритм кодування нам відомий, то ми можемо декодувати повідомлення.

Види кодування:

- Завадостійке кодування.
- Коди із заданою виправною здатністю.
- Систематичні коди.
- Коди з виявленням помилок: код із перевіркою на парність; код з прямим повторенням; коди з інверсним повторенням (інверсні коди); кореляційний код.
- Коди з виявленням і виправленням помилок: код Хемінга; циклічні коди; рекурентні коди.
- Кодування із стисненням інформації: код Шеннона-Фано; код Хаффмена; динамічне кодування методом Хаффмена; динамічне кодування методом FGK; динамічне кодування методом Віттера.

Характеристики кодів:

- довжина коду – кількість елементів (символів, знаків), які складають кодову комбінацію;
- кількість інформаційних елементів;
- кількість перевірочних елементів (для коректувальних роздільних кодів);

- алфавіт коду Q – множина символів (знаків), що різняться між собою, яка використовується для побудови кодових комбінацій коду (для двійкових кодів);
- потужність алфавіту коду – кількість символів (знаків) алфавіту (для двійкового коду);
- потужність коду – кількість дозволених кодових комбінацій, які використовуються для передачі повідомлень (для блокових роздільних кодів у загальному вигляді, зокрема для двійкових кодів);
- повна кількість кодових комбінацій N – кількість всіх можливих комбінацій для даного коду, яка підходить для блокових рівномірних кодів у загальному вигляді, зокрема для двійкового коду;
- надмірність коду, яка визначається:
 - для роздільних кодів;
 - та для нероздільних кодів: $R = 1 - (\log_2 N_0 / \log_2 N)$
- швидкість коду для роздільних кодів;
- вага кодової комбінації w – для двійкового коду визначається кількістю одиниць у кодовій комбінації;
- кодова відстань між двома кодовими комбінаціями однакової довжини визначається як кількість одноіменних елементів (розрядів) з різними значеннями символів (відстань Хеммінга);
- мінімальна кодова відстань – визначається для коду в цілому як мінімальне значення кодових відстаней між усіма парами кодових комбінацій, що належать до даного коду. Мінімальна кодова відстань визначає його здатність виявляти та виправляти помилки.

Класифікація кодів:

За потужністю алфавіту (кількістю q символів алфавіту) коди розділяють на **двійкові** ($q = 2$) та **недвійкові** ($q > 2$). Останні називають ще *багатопозиційними або багатоосновними*.

За коректувальною здатністю коди розділяють на *безнадмірні та надмірні*.

До безнадмірних кодів належать так звані *прості або первинні коди*, які використовують для первинного кодування джерел повідомлень. Ці коди не дозволяють виявляти і виправляти помилки. Це пояснюється тим, що такі коди використовують всі

можливі комбінації і будь-яка помилка призводить до появи дозволеної *кодової комбінації*.

До надмірних кодів належать коди, які дозволяють виявляти та/або виправляти помилки. У цих кодах використовується тільки визначена частина можливих комбінацій, що дозволяє при спотворенні елементів кодових комбінацій виявити або виправити їх. Кількість виявлених або виправлених помилок буде залежати від коректувальної здатності коду, тобто від його надмірності та особливостей побудови.

За способом побудови коди розділяють на *блокові та неперервні*. До першої групи належать коди, які є сукупністю кодових комбінацій, а до другої – коди, для яких кодування і декодування становлять неперервний у часі процес.

Блокові коди за кількістю елементів у кодових комбінаціях коду розділяють на рівномірні та нерівномірні. До першої групи належать коди, у яких всі комбінації, що складають код, мають однакову кількість елементів, а до другої – ті, у яких кодові комбінації коду можуть містити різну кількість елементів.

Блокові коди за використанням перевірочних елементів у кодових комбінаціях розділяють на роздільні та нероздільні. До першої групи належать коди, кодові комбінації яких містять інформаційні та перевірочні елементи, до другої – коди, у кодових комбінаціях яких не відокремлюються інформаційні та перевірочні елементи.

Блокові коди *за способом побудови перевірочних елементів у кодових комбінаціях* розділяють на лінійні (групові, систематичні) *та нелінійні* (несистематичні). До першої групи належать коди, у яких перевірочні елементи одержують як результат лінійних операцій над визначеними інформаційними елементами (для двійкових кодів за модулем 2), а до другої – коди, що будуються за іншими принципами.

Різновидом лінійних (систематичних) кодів є **циклічні коди**, для котрих характерною є така особливість: циклічний зсув будь-якої комбінації коду дає комбінацію, яка завжди належить до цього коду. При виборі коду для передачі даних керуються вимогами до вірогідності інформації, що передається, та швидкістю передачі даних, які визначаються, головним чином, характеристиками кодів.

Способи подання кодів

Залежно від застосовуваних методів кодування, використовують різні математичні моделі кодів, при цьому найбільш часто застосовується подання кодів у вигляді: кодових матриць; кодових дерев; многочленів; геометричних фігур і т.д.

2. Приклади бінарних кодів без надлишковості

У кодах без надлишку кількість інформації, яка припадає на один символ, максимальна. Однак, вони не можуть не тільки виправити, а часто і виявити помилку у реальних каналах зв'язку з перешкодами.

Згідно з (2.2) для передачі N повідомлень (кодових слів) бінарним кодом потрібна кількість розрядів (довжина коду):

$$n = \log_2 N$$

Запис кодових слів у вигляді двійкових чисел (2.3), у яких всі символи інформаційні, зветься простим двійковим кодом. У такому коді кількість інформації, що припадає на один символ, максимальна. Однак простий двійковий код не може не тільки виправити, але навіть виявити помилку.

Бінарними ненадлишковими кодами, які здатні виявляти деякі помилки, є так звані рефлексні (відображені) коди, у яких сусідні кодові слова відрізняються символом тільки в одному розряді. Відсутність такої властивості у прийнятих сусідніх кодових словах означає наявність помилки. *Серед рефлексних кодів найбільш розповсюдженим є код Грея.*

Перехід від простого двійкового коду (2.3) a_k , ($k = 1, 2, \dots, n$) до коду Грея $\hat{a}_k$ здійснюється за правилом:

$$\begin{aligned} \hat{a}_k &= a_k \\ \hat{a}_k &= a_k \oplus \hat{a}_{k-1}, (k = 1, 2, \dots, n) \end{aligned} \quad (2.11)$$

Старші розряди співпадають, а кожний наступний $\hat{a}_k$ дорівнює сумі за модулем 2 відповідного $\hat{a}_k$ і попереднього розрядів простого двійкового коду, зворотній перехід від коду Грея до простого двійкового відбувається за правилом:

$$\begin{aligned} \hat{a}_k &= a_k \\ a_k &= \hat{a}_{k-1} \oplus \hat{a}_k, (k = 1, 2, \dots, n) \end{aligned} \quad (2.12)$$

$$a_k = \overset{\wedge}{a_1} \oplus \overset{\wedge}{a_2} \oplus \dots \oplus \overset{\wedge}{a_k}$$

Старші розряди співпадають, а кожний наступний дорівнює сумі за модулем 2 всіх попередніх розрядів включно коду Грея. Практично при переведенні у код Грея [8, с.125] під двійковим числом записується таке саме число із зсувом праворуч на один розряд; при цьому молодший розряд зсуненого числа губиться. Наприклад переведення простого двійкового числа 11010011 у код Грея:

11010011

$$11010011 \Rightarrow [11010011]_2 = [10111010]_{\text{Грея}}$$

10111011

Практично при зворотньому переході до простого двійкового коду кожна цифра у коді Грея інвертується стільки разів скільки одиниць перед нею. Так для наведеного прикладу:

$$[10111010]_{\text{Грея}} = [1(0 \rightarrow 1)(1 \rightarrow 0)(1 \rightarrow \rightarrow 1)(1 \rightarrow \rightarrow \rightarrow 0)(0 \rightarrow \rightarrow \rightarrow \rightarrow 0)(1 \rightarrow \rightarrow \rightarrow \rightarrow 1)(0 \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow 1)]_2 = [11010011]_2$$

Бінарним кодом майже без надлишковості, який застосовується у телемеханіці і зв'язку для передачі текстових і цифрових повідомлень, є код Бодо. Це бінарний код довжиною $n = 5$, з загальним числом кодових слів:

$N = 2^5 = 32$ Ансамбль повідомлень коду Бодо – літери кириличної та латинської абеток, арабські цифри, розділові знаки («.», «.», «—», «:»), прогалина, службові слова. 32 кодових слів недостатньо для передачі всіх зазначених повідомлень, тому в апаратурі, що працює на коді Бодо, є три регістри. При цьому загальна кількість кодових слів майже потроюється і складає 86.

Якщо кодувати у коді Бодо кириличну абетку, то з урахуванням нерівномірності літер ентропія $H = 4,36 \frac{\text{Бит}}{\text{символ}}$. Максимальне значення ентропії (для рівномірної 32-літерної абетки) $H_{\max} = \log_2 32 = 5 \frac{\text{Бит}}{\text{символ}}$

Отже надлишковість коду Бодо:

$$r = 1 - \frac{H}{H_{\max}} = 1 - \frac{4,36}{5} = 0,128$$

Використовуються рівномірні коди, в яких букви алфавіту представляються комбінаціями з однакового числа елементів, і нерівномірні коди, складені з комбінацій різної довжини. Прикладом рівномірного коду може служити телетайпний код Бодо, а нерівномірного – азбука Морзе, які для деяких українських букв мають вигляд:

Буква	А	Б	В	Г	Д	Е
Код Бодо	10000	00110	01101	01010	11110	01000
Азбука Морзе	•–	–•••	--	–•–	–••	•

3. Основні теореми кодування

Якщо у інформаційній системі шуми малі, то яким би чином не кодувалися повідомлення, втрат інформації не буде і основною є проблема ефективності кодування, тобто такого у абетці з t символів інформації, що міститься у ансамблі з N повідомлень x_n і заданим розподілом $p_k = p(x_k), (k=1,2,...,N_0)$, щоб кількість інформації на один символ кодового слова була максимальною. Індивідуальна кількість інформації у кодовому слові (повідомленні) $I(x_k) = -\log_2 p_k$, середня кількість інформації у кодовому слові дорівнює ентропії джерела повідомлень:

$$H(N) = -\sum_{k=1}^{N_0} p_k \log p_k \quad (2.13)$$

а середня довжина кодового слова:

$$\bar{n} = \sum_{k=1}^N n_k p_k \quad (2.14)$$

залежить від способу кодування. Тому виникають питання принципового характеру: в яких межах знаходиться середня довжина кодових слів $\bar{n}$, які за цього індивідуальні довжини кодових слів n_k , як залежать величини $\bar{n}$ та n_k , від розподілу ймовірностей ансамблю кодових повідомлень, щоб забезпечувались зазначені у розділі 1 вимоги до кодових слів (взаємно – однозначна відповідність повідомленням, роздільність при декодуванні, мінімальні довжини)?

Індивідуальні довжини кодових слів, що задовольняють необхідним і достатнім умовам взаємно-однозначної відповідності повідомленням і роздільності (жодне кодове слово не є початком іншого більш довгого визначається теоремою (нерівністю) Крафта:

$$\sum_{k=1}^N m^{-n_k} \leq 1 \quad (2.15)$$

(Доведення. Нехай максимальна довжина слова у коді $n_{\max}$, тоді можлива кількість цих слів $m^{n_{\max}}$. Із зазначених $n_{\max}$ символів жодні перші n_k символів не повинні співпадати з словами довжиною n_k ($n = 1, 2, \dots, N$), отже кількість символів, не зайнятих словами довжиною n_k , дорівнює $n_{\max} - n_k$. Таким чином, загальна кількість $n_{\max}$ - символних слів, які можливо однозначно декодувати \

$$m^{n_{\max}-n_1} + m^{n_{\max}-n_2} + m^{n_{\max}-n_3} + \dots + m^{n_{\max}-n_k} \leq m^{n_{\max}}$$

Поділивши цю нерівність на $m^{n_{\max}}$ дістанемо (2.15).

Зв'язок між ентропією ансамбля повідомлень (2.15) і середньою довжиною слова (2.14) для системи без шумів визначається фундаментальною теоремою кодування (ФТК, Шеннон К.).

Для будь-якого ансамбля повідомлень з ентропією $h(n)$ існує код з основою m , для якого середня довжина кодового слова $\bar{n}$ задовольняє нерівностям:

$$\frac{H(N)}{\log_2 m} \leq \bar{n} \leq \frac{H(N)}{\log_2 m} + 1 \quad (2.16)$$

Доведення: Умова взаємно-однозначної відповідності кодових слів повідомленням:

$$N \leq m^{\bar{n}} \Rightarrow \log_2 m^{\bar{n}} \geq N \Rightarrow \bar{n} \geq \frac{\log_2 N}{\log_2 m} \Rightarrow (\log_2 N = H_{\max}(N) \phi H(N)) \Rightarrow \bar{n} \geq \frac{H(N)}{\log_2 m}$$

Отже, мінімальне ціле число n для найефективнішого коду (якщо він існує) задовольняє (2.16).

Доведемо, що такий код існує. Задамо n_k нерівностями:

$$-\frac{\log_2 P_k}{\log_2 m} \leq n_k \leq -\frac{\log_2 P_k}{\log_2 m} + 1$$

(у такому випадку для середніх значень виконується (2.16):

$-n_k \log_2 m \leq \log_2 p_k \Rightarrow \log_2 m^{-n_k} \leq \log_2 p_k \Rightarrow m^{-n_k} \leq p_k \Rightarrow \sum_{k=1}^N m^{-n_k} = \sum_k p_k = 1$
 , тобто умова (2.16) задовольняє нерівності Крафта (2.15)).

Якщо відбувається не поелементне кодування, а поблокове: кодове слово ставиться у відповідність не одному повідомленню, а блоку з M повідомлень, то ФТК має вигляд:

$$\frac{H(N)}{\log_2 m} \leq \bar{n} \leq \frac{H(N)}{\log_2 m} + \frac{1}{M}$$

За великої довжини $n \gg 1$ ймовірність типових (найймовірніших) кодових слів:

$$P_{\max} = 2^{-nH(m)}$$

де $H(m)$ – ентропія абетки коду.

Доведення. У кодовому слові довжиною $n \gg 1$ кількість символів коду з ймовірністю p_i ($i=1,2,\dots,m$) за законом великих чисел дорівнює np_i і ймовірність такого типового (найймовірнішого) слова за незалежності його символів:

$$P_{\max} = P_1^{n p_1} \cdot P_2^{n p_2} \wedge P_m^{n p_m} \Rightarrow -\log_2 P_m = n(-P_1 \log_2 P_1 - P_2 \log_2 P_2 - \wedge - P_m \log_2 P_m) = nH(m) \Rightarrow P_{\max} = 2^{-nH(m)}$$

Наслідок. Всі типові слова практично рівноймовірні і їх кількість:

$$N_{\min} = \frac{1}{P_{\max}} = 2^{nH(m)} \quad (2.17)$$

За наявності у каналі зв'язку шумів, основна вимога до кодування – завадостійкість, тобто можливість зведення ймовірності помилок (спотворень символів) до прийнятного рівня.

Завадостійкість коду і швидкість кодування (2.10) (кількість інформації, що кодується за одиницю часу) у визначається теоремами Шеннона (прямою і оберненою) [8, с.92].

Пряма теорема кодування (Шеннон)

За будь-якої швидкості кодування, меншої ніж пропускна здатність каналу зв'язку, існує код, що забезпечує як завгодно малу ймовірність помилки.

Доведення. Ймовірність помилки, це ймовірність появи на приймальному боці каналу зв'язку кодового слова, що не передавалося.

Для двосимвольного коду ($n \gg 1$) з ентропією $H_n = \max_{\text{по-кодах}} H(m)$, що відповідає пропускній здатності каналу зв'язку $c = \frac{H_n - H(Ш)}{T}$ ентропія на виході каналу $H_n - H(Ш) = cT$ і за теоремою асимптотичної рівноймовірності кількість прийнятих типових слів: $2^{n(H_n - H(Ш))} = 2^{ncT}$

За швидкості кодування $\tau = \frac{H(m)}{T}$ кількість всіх типових слів $2^{nH(m)} = 2^{nT}$.

Отже ймовірність появи серед прийнятих кодового слова, що не передавалось – ймовірність помилки:

$$P_{\max} = \frac{2^{H\tau T}}{2^{ncT}} = 2^{-nT(c-\tau)}$$

Якщо $\gamma < c$,

то

$$\lim_{n \rightarrow \infty} P_{\max} = \lim_{n \rightarrow \infty} 2^{-nT(c-\tau)} = 0$$

Отже, при кодуванні довгими словами ймовірність помилки може бути як завгодно малою.

4. Оптимальне кодування за відсутності шумів.

Код Шеннона-Фано

Наведені вище основні теореми кодування є теоремами існування: вони лише доводять існування оптимальних (ефективних, завадостійких) кодів, але не дають способів їх побудови.

Нижче розглядаються питання ефективного бінарного кодування за відсутності шумів. Для таких систем оптимальними вважаються майже безнадлишкові коди з мінімальною середньою довжиною слів (2.13).

Властивості бінарних оптимальних кодів:

1. Довжина кодового слова повинна бути обернено пропорційна його ймовірності (тобто ймовірності відповідного повідомлення):

$$(з (2.5) і (2.14) випливає \sum_{k=1}^N n_k p_k \approx H(m) \approx H_{\max}(m) = 1 \Rightarrow n_k \approx \frac{1}{p_k})$$

2. Символи абетки коду (0 та 1) повинні бути взаємонезалежними і рівноймовірними (це відповідає умові $H(m) \approx H_{\max}(m)$).

3. Кодове слово не може бути початком іншого, більш довгого (можливість відокремлення без додаткових символів).

4. Нерівномірність.

Прикладом наближення до ефективного бінарного коду є алгоритм Шеннона-Фано:

1. Множина повідомлень, що кодуються, розташовується у порядку спадання ймовірностей.

2. Зазначена множина розбивається на дві підгрупи таким чином, щоб суми ймовірностей повідомлень в обох групах були по можливості однакові.

3. Першій підгрупі присвоюється символ 0, другій підгрупі – символ 1. Так отримується перша цифра коду.

4. Кожна з отриманих підгруп знову розбивається на дві групи і першим групам кожної підгрупи знову присвоюється 0, а другим – 1. Так отримується друга цифра коду.

5. Кожна з отриманих чотирьох груп знову розбивається на дві частини, і процес триває доки не будуть закодовані всі повідомлення.

5. Принципи завадостійкого кодування

За наявності шумів у каналі, одна з основних вимог до кодування – завадостійкість, тобто, можливість доведення ймовірності помилок (спотворення символів) до прийнятого рівня. Можливості коду виявляти і виправляти помилки пов'язані з внесенням надлишкових (неінформаційних) символів у кодовому слові. При побудові коду для передачі інформації використовується лише частина кодових слів (дозволені кодові комбінації), які відрізняються одне від одного більш, ніж одним символом. Всі інші кодові слова не використовуються і належать до числа заборонених. При цьому, помилка в одному символі призводить до заміни дозволеного слова забороненим, що виявляє помилку. При достатній відмінності дозволених кодових слів (кодовій відстані) таким чином можна виявляти двократну, трикратну помилки, оскільки вони призводять до заборонених кодових слів.

Приклад. Для рівномірного бінарного коду, що кодує 32 повідомлення $N = 2^k = 32; k = 5$ і виправляє всі одно- і двократні помилки ($S = 2$):

$$\frac{2^n}{32} \geq C_n^0 + C_n^1 + C_n^2 = 1 + n + \frac{n(n-1)}{2} = 1 + \frac{n(n+1)}{2} \Rightarrow n \geq 12$$

$$(n=11 \Rightarrow 2^6 = 64 < 1 + \frac{11 \cdot 12}{2} = 67; n=12 = 128 > 1 + \frac{12 \cdot 13}{2} = 79)$$

Наведена нерівність, як і подібні до неї, визначає лише межу довжини коду (і, відповідно, кількості перевірочних символів), яка далеко не в усіх випадках може бути практично реалізована.

До найбільш поширених кодів з виявленням та виправленням помилок відносять код Хемінга, циклічні та рекурентні коди.

Код Хемінга відноситься до систематичних кодів, в яких з n символів, які утворюють комбінацію, n_0 символів є інформаційними, а останні $k=n-n_0$ з надлишковими (контрольними), призначеними для перевірки (контрольні символи у всіх комбінаціях займають однакові позиції). Коди Хемінга дозволяють виправити всі одиничні помилки (при кодовій відстані $d=3$) і визначити всі подвійні помилки (при $d=4$), але не виправляти їх.

Зв'язок між кількістю інформаційних та контрольних символів в коді Хемінга знаходять на основі таких міркувань. При передачі комбінації по каналу з шумами може бути спотворений довільний з n символів коду, або комбінація може бути передана без спотворень. Таким чином може бути $n+1$ варіантів спотворення (включаючи передачу без спотворення). Використовуючи контрольні символи, необхідно перевірити всі $n+1$ варіантів. За допомогою контрольних символів k можна описати 2^k подій. Для цього повинна бути використана умова:

$$2^k \geq n+1 = n_0 + k + 1.$$

В таблиці 2.2 подана залежність між k і n_0 , яка отримана з цієї нерівності, де k – число контрольних символів в коді Хемінга, n_0 – інформаційних символів.

Таблиця 2.2 – Розміщення контрольних символів в комбінаціях коду Хемінга

n_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
k	2	3	3	3	4	4	4	4	4	4	4	5	5	5	5

В коді Хемінга контрольні символи розташовують на місцях, кратних степеню числа 2, тобто на позиціях 1, 2, 4, 8 і т. п. Інформаційні символи розташовують на місцях, що залишилися. Наприклад, для семиелементної закодованої комбінації можна записати:

k_1	k_2	a_{04}	k_3	a_{03}	a_{02}	a_{01}
a_1	a_2	a_3	a_4	a_5	a_6	a_7

Символи коду Хемінга, які обведені прямокутниками, є *контрольними*, останні – *інформаційні*, де a_3 – старший (четвертий) розряд вихідної кодової комбінації двійкового коду, який необхідно кодувати, a_7 – молодший (перший) розряд. Після розташування на відповідних місцях кодової комбінації контрольних та інформаційних символів в коді Хемінга складають спеціальні перевірочні рівняння, які використовують для визначення наявності спотворень і їх виправлення. З перевірних рівнянь і отримують контрольні символи при кодуванні вихідної кодової комбінації двійкового коду. Для визначення контрольних символів необхідно використати такий алгоритм.

1. Всі символи коду Хемінга з номерами розрядів розташовують в порядку збільшення номерів і під ними записують номери розрядів в двійковому коді:

a_1	a_2	a_3	a_4	a_5	a_6	a_7
0001	0010	0011	0100	0101	0110	0111

2. Перше перевірочне рівняння складають як суму за mod 2 всіх розрядів, в номерах яких в молодшому розряді 2^0 стоїть одиниця: $S_1 = a_1 \oplus a_2 \oplus a_5 \oplus a_7$.

Друге перевірочне рівняння складають як суму за mod 2 всіх розрядів, в номерах яких стоїть одиниця на другому місці відповідного двійкового еквівалента (2^1): $S_2 = a_2 \oplus a_3 \oplus a_6 \oplus a_7$.

Третє перевірочне рівняння складають як суму за mod 2 всіх розрядів, в номерах яких стоїть одиниця на третьому місці (2^2): $S_3 = a_4 \oplus a_5 \oplus a_6 \oplus a_7$.

Аналогічно утворюються і інші перевірочні суми (при більшій кількості інформаційних і контрольних символів, відповідно).

Як видно з наведених рівнянь, в кожному перевірочному суму входить тільки один невизначений контрольний символ k_i (a_1 , a_2 , a_4 , відповідно), а всі інші інформаційні символи відомі.

Всі перевірочні рівняння за умовою Хемінга повинні дорівнювати 0 при підсумовуванні за mod 2. З цієї умови і знаходять контрольні символи.

Наприклад, необхідно передати інформаційну кодову комбінацію:

a_1	a_2	a_3	a_4
1	1	0	0

з числом розрядів $n_0 = 4$

З формули $2^k \cdot n + 1 = n_0 + 1 = n_0 + 1$ визначимо, що $k = 3$. Запишемо перевірочні суми:

$$S_1 = a_1 + a_3 + a_5 + a_7;$$

$$S_2 = a_2 + a_3 + a_6 + a_7;$$

$$S_3 = a_4 + a_5 + a_6 + a_7;$$

Підставимо у рівняння значення відомих інформаційних символів. З умови рівності нулю всіх перевірочних сум визначаємо відповідно контрольні символи. З першого рівняння $a_0 = 0$, з другого $a_1 = 1$, з третього $a_4 = 1$.

Відповідно буде передана така комбінація:

a_1	a_2	a_3	a_4	a_5	a_6	a_7
0	1	1	1	1	0	0

яка є комбінацією коду Хемінга.

Алгоритм декодування коду Хемінга

На приймальному пристрої комбінація коду Хемінга декодується – визначається наявність спотворення прийнятої кодової комбінації і, якщо один символ комбінації спотворений, він автоматично спочатку визначається, а потім виправляється. Визначення і виправлення спотвореного символу здійснюється на основі перевірних рівнянь. При правильному прийомі всі суми повинні дорівнювати нулю. В разі спотворення двійкове число, яке є результатом цих сум (синдром помилки), переводять в десяткове число, яке вказує номер спотвореного символу, який в подальшому виправляється шляхом інвертування.

Наприклад, при прийомі закодованої вище неспотвореної кодової комбінації Хемінга:

a_1	a_2	a_3	a_4	a_5	a_6	a_7
0	1	1	1	1	0	0

значення всіх символів комбінації підставляють на відповідні місця у перевірних сумах:

$$S_1 = a_1 + a_3 + a_5 + a_7 = 0;$$

$$S_2 = a_2 + a_3 + a_6 + a_7 = 0;$$

$$S_3 = a_4 + a_5 + a_6 + a_7 = 0;$$

Ознака правильно прийнятої комбінації – рівність нулю всіх сум.

Припустимо, що шостий символ кодової комбінації спотворений, тобто замість комбінації 0111100 буде прийнята:

a_1	a_2	a_3	a_4	a_6	a_7	a_8
0	1	1	1	1	1	0

Підставимо значення прийнятих символів у перевірочні рівняння.

Одержимо:

$$S_1 = a_1 + a_3 + a_5 + a_7 = 0;$$

$$S_2 = a_2 + a_3 + a_6 + a_7 = 0;$$

$$S_3 = a_4 + a_5 + a_6 + a_7 = 0;$$

Переведемо двійкове число (синдром) $1^2 1^2 1^0$ в десяткове. Одержимо десяткове число 6, яке і вказує на номер спотвореного символу.

Код з прямим повторенням

Існує ряд кодів, в яких для збільшення завадозахищеності до інформаційної кодової комбінації, яка складається з n_0 інформаційних символів двійкового безнадлишкового коду, додається k контрольних символів, причому $k=n_0$. Таким чином, загальна кількість елементів в кодовій комбінації дорівнює $2n_0$. Тому такі коди називають *кодами з подвоєнням числа елементів*.

В залежності від способу утворення контрольних символів такі коди класифікують на:

- а) коди з прямим подвоєнням;
- б) коди з інверсним подвоєнням;
- в) кореляційні коди.

Розглянемо більш детально схеми, що реалізують вказані вище коди.

Кодер коду з прямим повторенням – це пристрій, який автоматично добавляє до інформаційної кодової комбінації

двійкового коду таку ж саму (повторює її), збільшуючи загальну кількість двійкових символів вдвічі.

Коди з інверсним повторенням (інверсні коди)

Коди з інверсним повторенням інформаційної двійкової кодової комбінації утворюються за таким правилом: якщо кількість одиниць в інформаційній комбінації парна, то до неї автоматично додається така сама комбінація (пряме повторення) і якщо кількість одиниць в інформаційній комбінації непарна, до неї додається інвертована інформаційна комбінація. Наприклад, якщо інформаційна кодова комбінація має вигляд 1100, то до неї додається така сама комбінація 1100 і вцілому для подальшого перетворення надійде комбінація 11001100.

Якщо інформаційна комбінація має вигляд 1110, то до неї буде додана комбінація 0001 і комбінація інверсного коду буде мати вигляд 11100001.

Кодером коду з інверсним повторенням називається електронний пристрій, який до інформаційної n_0 – розрядної комбінації двійкового коду автоматично додає контрольну комбінацію з $k=n_0$ символів за таким алгоритмом: якщо кількість одиниць інформаційного коду парна, то до неї додається така ж сама комбінація, якщо непарна, до неї додається інвертована інформаційна комбінація.

Кореляційний код

Кодером кореляційного коду [8, с. 136] є електронний пристрій, який при надходженні на його входи n_0 інформаційних символів двійкового коду автоматично додає до кожного інформаційного символу по одному контрольному "0" чи "1" і значення якого утворюється за правилом: інформаційний нуль перетворюється в 01, інформаційна одиниця перетворюється в 10. Таким чином загальна кількість символів кодової комбінації кореляційного коду $n = n_0 + k = 2n_0$. Тому цей код відносять до коду з подвоєнням числа елементів.

Коди з виявленням і виправленням помилок

До найбільш поширених кодів з виявленням та виправленням помилок відносять код Хемінга, циклічні та рекурентні коди.

Рекурентні коди

Циклічні коди дозволяють виявляти і виправляти як одиночні і подвійні помилки, так і пакети помилок. Однак, практичне застосування цих кодів для виправлення пакетів помилок не є вдалою ідеєю тому, що при не дуже великій надмірності довжина кодових комбінацій значно перевищує довжину пакетів. У цьому відношенні більш зручними є рекурентні коди.

Рекурентні (неперервні) коди відрізняються тим, що в них операції кодування і декодування робляться неперервно над послідовністю символів без розподілу їх на блоки.

Ці коди призначені, в основному, для виправлення пакетів помилок. Формування перевірних символів здійснюється шляхом додавання двох або більше інформаційних символів, зміщених один відносно одного на визначену відстань t , звану кроком додавання.

Кодування із стисненням інформації

Застосування стиснення даних дозволяє більш ефективно використовувати ємність дискової пам'яті. Не менш корисне застосування стиснення при передачі інформації в будь-яких системах зв'язку. У останньому випадку з'являється можливість передавати значно меншу (як правило, у декілька разів) кількість даних і, отже, необхідні значно менші ресурси пропускну здатності каналів для передачі тієї ж самої інформації.

6. Шифрування інформації

Шифр Цезаря

Шифр Цезаря [13, с.14] (Caesar, 100-44 pp. до н.е.) реалізує кодування фрази шляхом «зміщення» усіх букв фрази на певне число kk (у оригінальному шифрі Цезаря це число kk дорівнювало 3). Якщо буква шифрованої фрази має в алфавіті позицію jj , то вона в «шифровці» замінюватиметься буквою, що знаходиться в алфавіті на позиції $jj + kk$.

Нехай $kk = 3$ і фразою для шифрування буде «i remember that september». Використовуватимемо латинські букви із стандартним проходженням букв в алфавіті.

1	i		r	e	m	e	m	b	e	r		t	h	a	t	
2	9	0	18	5	13	5	13	2	5	18	0	20	8	1	20	0
3	12	3	21	8	16	8	16	5	8	21	3	23	11	4	23	3
4	12	3	21	8	16	8	16	5	8	21	3	23	11	4	23	3
5	l	c	u	h	p	h	p	e	h	u	c	w	k	d	w	c

1	s	e	p	t	e	m	b	e	r
2	19	5	16	20	5	13	2	5	18
3	22	8	19	23	8	16	5	8	21
4	22	8	19	23	8	16	5	8	21
5	v	h	s	w	h	p	e	h	u

Результати шифрування вказаної вище фрази показані нижче в таблиці:

Пояснення до таблиці:

1-й рядок – фраза для шифрування;

2-й рядок – номери букв фрази для шифрування в латинському алфавіті;

3-й рядок – номери букв фрази для шифрування, збільшені на 3;

4-й рядок – результат «ділення по модулю 26» чисел 3-го рядка;

5-й рядок – зашифрована фраза.

Шифр Віженера

Шифр Віженера [13, с. 21] реалізує кодування фрази шляхом «індивідуального зміщення» букв, причому величина зміщень визначається номерами (положенням) букв в ключовому слові (фразі). Візьмемо, наприклад, ключове слово «leonid» (латиниця) і кодовану фразу «I remember that September». Букви ключового слова мають наступні номери в латинському алфавіті: 12,5,15,14,9,4. Шифрування по Віженеру полягає в «зсуві» першої букви кодованої фрази на 12 позицій, тобто в заміні букви «і» (9-а позиція) на букву «u», що знаходиться в $9+12=21$ -й позиції, в заміні пропуску (другої букви кодованої фрази, 0-а позиція) на букву «е», що знаходиться в $0+5=5$ -й позиції і так далі.

При «вичерпанні» букв ключового слова, останнє використовується знову і знову до тих пір, поки не будуть закодовані всі букви кодованої фрази. Результати шифрування вказаної вище фрази показані нижче в таблиці:

1	i		r	e	m	e	m	b	e	r		t	h	a	t	
2	9	0	18	5	13	5	13	2	5	18	0	20	8	1	20	0
3	l	e	o	n	i	d	l	e	o	n	i	d	l	e	o	n
4	12	5	15	14	9	4	12	5	15	14	9	4	12	5	15	14
5	21	5	33	19	22	9	25	7	20	32	9	24	20	6	35	14
6	21	5	7	19	22	9	25	7	20	6	9	24	20	6	9	14
7	u	e	g	s	v	i	y	g	t	f	i	x	t	f	i	n

1	s	e	p	t	e	m	b	e	r
2	19	5	16	20	5	13	2	5	18
3	i	d	l	e	o	n	i	d	l
4	9	4	12	5	15	14	9	4	12
5	28	9	28	25	20	27	11	9	30
6	2	9	2	25	20	1	11	9	4
7	b	i	b	y	t	a	k	i	d

Пояснення до таблиці.

1-й рядок – фраза для шифрування;

2-й рядок – номери букв фрази для шифрування в латинському алфавіті;

3-й рядок – ключове слово з довжиною рівній довжині фрази;

4-й рядок – номери букв ключового слова в алфавіті;

5-й рядок – сума номерів 2-го і 4-го рядків у відповідних стовпцях;

6-й рядок – результат «ділення по модулю 26» чисел 5-го рядка;

7-й рядок – зашифрована фраза.

Таким чином, «шифровка» матиме вигляд «uefsviygteixtfhnaiaiytkic».

Декодування «шифровки» проводимо «аналогічно» кодуванню фрази.

Шифр підстановки (квадрат Полібія)

Алфавіт початкового повідомлення записати у вигляді матриці. Кожну букву кодувати парою чисел (рядок, стовець).

Наприклад, візьмемо квадрат 5x5, стовпці і рядки якого нумерувалися від 1 до 5. У кожен клітинку цього квадрату записувалася одна буква [13, с.12].

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I,J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

Зашифруємо фразу Декарта «Cogito, ergosum» («Я мислю, отже, існую»). Тоді результатом буде наступна послідовність чисел.

13	34	22	24	44	34	15	42	22	34	43	45	32
C	O	G	I	T	O	E	R	G	O	S	U	M

Штриховий код

Штриховий код [8, с.146] – це товарний знак, який наноситься на товар або його упаковку у вигляді штрихового або цифрового символу, зчитуваного сканером.

Штриховий код є одним із засобів системи автоматичної ідентифікації товару, до якої також належать засоби цифрової, магнітної, радіочастотної, звукової та візуальної ідентифікації (магнітна картка, радіочастотна бірка і т. д.). Його головна перевага перед іншими засобами автоматичної ідентифікації полягає в можливості оперативно передавати інформацію про товар за системою електронного зв'язку, тобто ШК є ефективним засобом телекомунікації.

Штриховий код призначений:

- для оперативної ідентифікації товару і виробника;
- проведення торгових операцій «без паперів»: ШК скорочує витрати на діловодство від 15 до 0,5-3% вартості товару;
- автоматизованого обліку та контролю товарних запасів;
- оперативного управління процесом руху товару: відвантаження, транспортування і складування товарів (продуктивність праці по забезпеченню руху товару підвищується на 30%, у деяких випадках – до 80%);
- інформаційного забезпечення маркетингових досліджень.

Штрихові коди застосовуються в промисловому виробництві, оптовій та роздрібній торгівлі, складському господарстві, транспорті, банківській справі, митному контролю, страхуванні, охороні здоров'я, видавництві та ін.

Кожному виду, різновиду, модифікації товару присвоюється індивідуальний товарний номер (позначається штрих-кодом). При зміні споживчих властивостей товару (складу, розміру, маси, способу упаковки, комплектності і т. д.) його ШК змінюється. тільки ринкове коливання ціни на товар не тягне за собою зміни його товарного номеру (коду).

Існують декілька видів кодів, серед яких найбільш поширеними є європейські коди типу EAN і американські типу UPC.

Коди EAN залежно від числа знаків символів підрозділяють на EAN-8, EAN-13 і EAN-14.

За структурою кодів розрізняють ШК:

- дискретні – знаки розділені міжзнаковими інтервалами;
- безперервні – знаки-роздільники відсутні;
- двонаправлені – можна зчитувати у двох напрямках – зліва направо і зправа наліво.

Найпоширенішими є такі штрихові коди:

- код «2 з 5 Industrial»;
- код «2 з 5 Matrix»;
- код «2 з 5 з чергуванням» або код ITF (Interleaved Twenty Five);
- код 39;
- код 93;
- код Codabar;
- код 128;
- код EAN (European Article Number);
- код UPC (Uniform Product Code);
- код 49.

Структура штрихового коду визначає основні характеристики його символіки:

- кількість знаків у наборі;
- контролеприднадність;
- щільність штрихового коду.

Широкого розповсюдження набуло кодування усіх товарів, незалежно від їхньої якості і престижності фірм-виготовлювачів. Крім того, самі ШК стали об'єктом фальсифікації. У цьому зв'язку досить важливо вміти розпізнавати цей вид інформаційної фальсифікації.

Візьмемо, наприклад, цифровий код: 4606453849072. Перші дві цифри (460) означають країну походження (виробника або

продавця) продукту, наступні 4 або 5 залежно від довжини коду країни (6453) – підприємство-виробник, ще п'ять (84907) – найменування товару, його споживчі властивості, розміри, масу, колір. Остання цифра (2) контрольна, що використовується для перевірки правильності зчитування штрихів сканером.

Тринадцятизначний номер штрихового коду EAN–13 включає:
перші 3 знаки – код банку даних організації, що зареєструвала підприємство-виробник і товар (видала товарний номер). Необхідно враховувати, що він не завжди збігається за місцем походження (країною виготовлення) товару, так як фірма може бути зареєстрована не у вітчизняному банку даних, а в закордонному;

4 знаки – код підприємства, що виробляє або реалізує товар;

5 знаків – код товару. Підприємство самостійно визначає кодовану інформацію (з урахуванням споживчих властивостей товару, упаковки, маси і т. д.) і доводить її до торгового партнера, який повинен отримати інформацію про ШК мінімум за 3 тижні до поставки самого товару (це правило діє і в разі зміни ШК);

1 знак (останній) – контрольна цифра розраховується шляхом найпростіших арифметичних операцій і призначена для перевірки правильності зчитування (сканування) ШК.

Для коду товару:

1 цифра: найменування товару,

2 цифра: споживчі властивості,

3 цифра: розміри, маса,

4 цифра: інгредієнти,

5 цифра: колір.

Нерідко на товарі можна побачити напис, наприклад, «Зроблено в Німеччині», а код, нанесений на етикетку, цій країні не відповідає. Причин цього може бути декілька:

✓ Фірма була зареєстрована і отримала код не в своїй країні, а в тій, куди направлений основний експорт продукції;

✓ Товар був виготовлений на дочірньому підприємстві, розташованому в іншій країні;

✓ Товар був виготовлений в одній країні, але за ліцензією фірми з іншої країни;

✓ Засновниками підприємства є кілька фірм з різних держав.

Використовуючи штрих-код, можна, з відомим ступенем достовірності, судити про справжність товару або ж встановити наявність фальсифікації продукції.

Це може бути зроблено за допомогою наявного в штрих-кодi контрольного знака (остання цифра штрих-коду).

Приклад: Розглянемо методику такого аналізу на прикладі штрих-коду типу EAN-13 (код 4820024700016).

Для такого аналізу слід провести наступні обчислення:

1. Скласти цифри, що стоять на парних місцях: $8 + 0 + 2 + 7 + 0 + 1 = 18$.

2. Отриману суму помножити на 3: $18 \cdot 3 = 54$.

3. Скласти цифри, що стоять на непарних місцях, без контрольної цифри: $4 + 2 + 0 + 4 + 0 + 0 = 10$.

4. Скласти числа, зазначені в пунктах 2 і 3: $54 + 10 = 64$.

5. Відкинути десятки: отримаємо 4.

6. З 10 відняти отримане в пункті 5: $10 - 4 = 6$.

Якщо отримана після розрахунку цифра не співпадає з контрольною цифрою в штрих-кодi, це означає, що товар вироблений незаконно, а якщо збігається – товар справжній.

Питання для самоконтролю

1. Що таке код, кодування інформації?
2. Що таке шифрування, дешифрування?
3. Назвіть вимоги до кодів.
4. Назвіть найбільш розповсюджені рефлексні коди.
5. Опишіть процес кодування Бодо.
6. В чому відмінність підходів Шеннона та Хартлі до проблеми оцінювання кількості інформації?
7. Наведіть властивості бінарних оптимальних кодів.
8. Дайте визначення кодування та декодування Хемінга.

Тема 3. Теорія автоматів

План

1. Основні поняття теорії автоматів
2. Мікропроцесорні цифрові автомати і їх структури
3. Визначення абстрактного автомата. Алфавіти входу, виходу, станів. Функції виходів і переходів
4. Завдання автоматів. Таблиці і матриці переходів та виходів. Об'єднана таблиця

1. Основні поняття теорії автоматів

Термін «автомат» використовується в двох аспектах. З одного боку, автомат – пристрій, що виконує деякі функції без участі людини (наприклад, ЕОМ). З іншого боку, термін «автомат» – математичне поняття і позначає математичну модель реальних технічних процесів.

Автомат – як математична модель, описує реальні технічні автомати пристроїв.

Теорія автоматів – це абстрактний опис технічних пристроїв, соціально-економічних, біологічних, обчислювальних і інших динамічних процесів або опис програм та алгоритмів; це створення моделі автомата, в основі моделей лежить припущення, що автомати працюють дискретним чином; знаходяться перед і після кроку у встановленому стані, і за кожний крок сприймають деякий вхід або вихід.

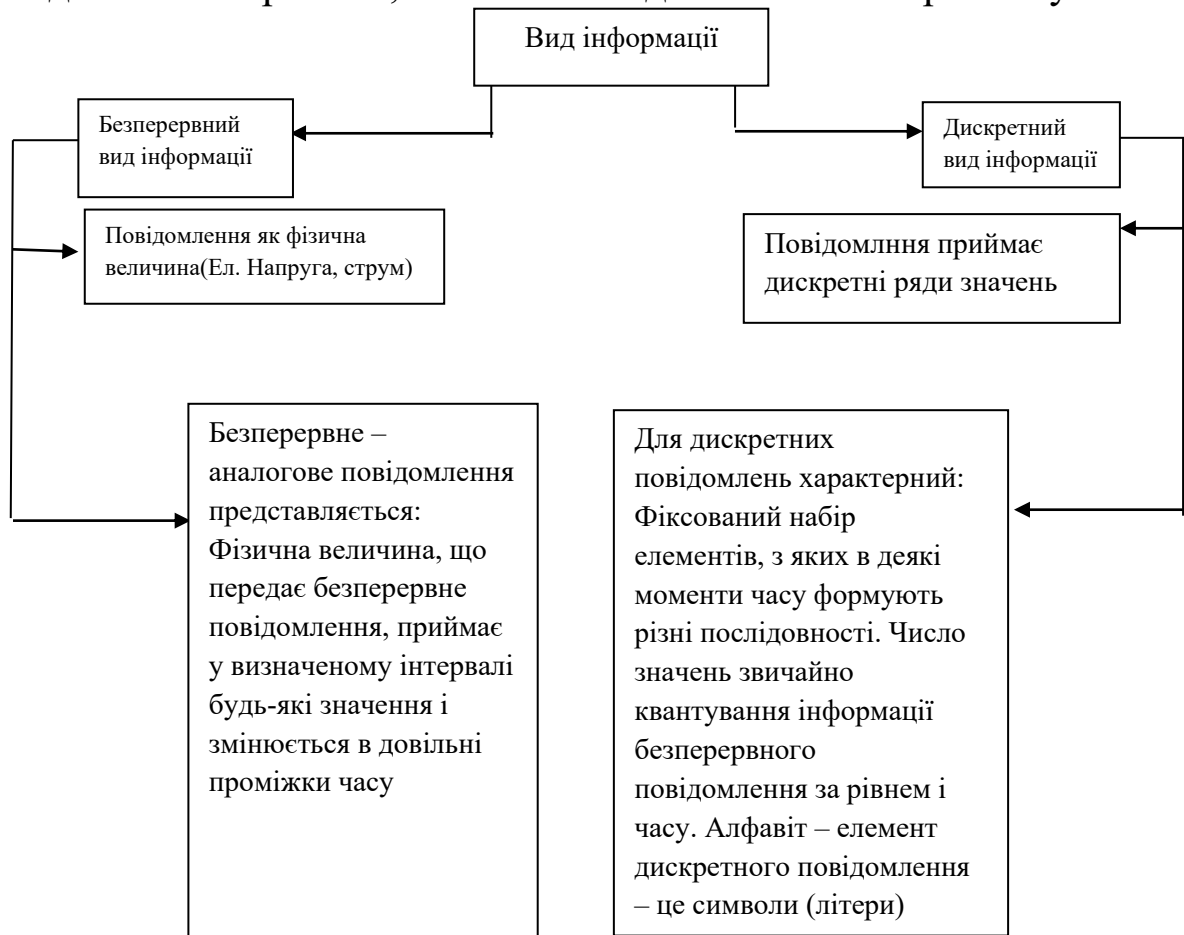
У загальному випадку автомат представляється як «чорний ящик» і повністю описується сукупністю наступних шести об'єктів:

- 1) множина вхідних сигналів X ;
- 2) безліч вихідних сигналів Y ;
- 3) безліч станів автомата A ;
- 4) початковий стан автомата, a_0 як елемент множини A ;
- 5) функція переходу з одного стану в інший $f(a, x)$;
- 6) функція виходів автомата.

Кінцевий (скінченний) автомат особливий різновид автомату – абстракції, що використовується для описання шляху зміни стану об'єкта в залежності від поточного стану та інформації отриманої ззовні [Вікепедія].

Цифровий автомат – пристрій, призначений для перетворення цифрової інформації.

Поняття “стан автомата” використовується для опису систем, виходи яких залежать не тільки від вхідних сигналів у даний момент часу, але і від деякої передісторії, – сигналів, які надійшли на входи системи раніше, тобто стан – деяка пам'ять про минуле.



Автомати, в яких послідовність сигналів, вироблюваних на його виходах, однозначно визначаються вхідними послідовностями, називаються детермінованими, на відміну від ймовірнісних автоматів, які виробляють випадкові послідовності вихідних сигналів.

Детерміновані автомати, в свою чергу, можна розділити на три **типи**, що відрізняються один від одного в функціональному відношенні.

Цифрові автомати ділять на два класи. У *синхронних автоматах* моменти часу, в які фіксуються стани автомата, задаються спеціальним пристроєм – генератором синхроімпульсів.

В *асинхронних автоматах* моменти переходу автомата з одного стану в інший заздалегідь не визначені й залежать від якихось подій.



Типи автоматів в залежності від особливостей вхідної стрічки і пам'яті:

- скінченні автомати;
- автомати з магазинною пам'яттю;
- лінійно-обмежені автомати;
- машини Т'юрінга.

Автомати останніх трьох типів зазвичай поєднують у клас нескінченних автоматів.

Функціональна модель автомата містить інформацію про те, як автомат працює.

Структурна модель автомата повинна показувати, як автомат влаштований, тобто, з яких елементів він складається, і як ці елементи пов'язані між собою.

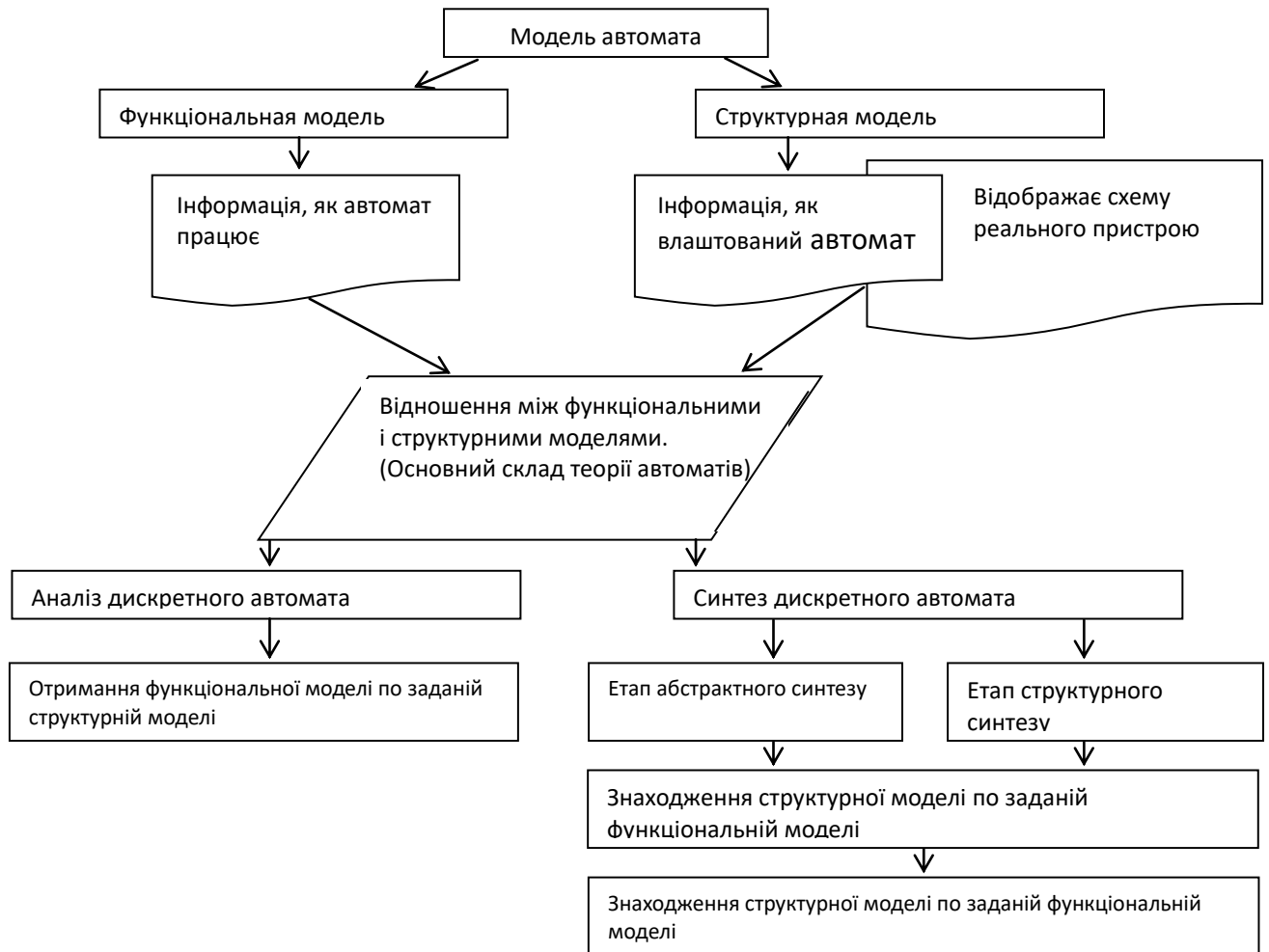
Структурна модель автомата відображає схему реального пристрою.

Основний зміст теорії автоматів становить дослідження відносин між функціональними та структурними моделями.

Синтез схем автоматів ділиться на два етапи:

- етап абстрактного синтезу, в процесі якого виявляється взаємодія елементів (словесній формі);
- обсяг пам'яті автомата (з стандартних форм).

Склад і опис автомата:



Всі цифрові автомати класифікуються за низкою ознак:

- 1) За законом функціонування цифрового автомата (автомат підрозділяється на 2 класи – 1 роду і 2 роду (Мілі та Мура) відмінною особливістю яких є залежність від вхідних сигналів).
- 2) За множиною вхідних і вихідних сигналів, а також множинами станів (кінцеві автомати і нескінченні автомати).
- 3) За обсягом пам'яті (з пам'яттю – послідовні і без пам'яті – схеми).
- 4) За ступенем розкриття структури (абстрактні та структурні).
- 5) За відношенням між автоматами (підавтомат і надавтомат).
- 6) За повнотою використовуваних переходів з одного стану в інший (повністю і частково визначені).
- 7) За стабільністю прямування вихідних сигналів (синхронні, асинхронні).

Цифровий автомат складається:

- пристрою запам'ятовування,
- операційного пристрою,
- пристрою керування.

2. Мікропроцесорні цифрові автомати і їх структури

Мікропроцесорний автомат має стандартну архітектуру, полегшує проектування спеціалізованих ІС (інформаційних систем) і дозволяє в кілька разів зменшити час від початку розробки до виходу кінцевого продукту.

Етапи процесу розробки:

- опис (на рівні блоків) алгоритму;
- створення необхідної системи команд;
- підключення до мікропрограмного автомату необхідних бібліотек;
- написання мікропрограм;
- налагодження мікропрограм.

Багатоблокові автомати це автомати, в яких керуючий автомат управляє блоками об'єктів управління, грубо кажучи, об'єкт управління не один.

3. Визначення абстрактного автомата. Алфавіти входу, виходу, станів. Функції виходів і переходів

Цифровий автомат описується математичною моделлю (абстрактним автоматом) і законом його функціонування.

Абстрактний автомат – це модель цифрового автомата, яка характеризується алфавітом входу, алфавітом виходу, алфавітом станів, функціями переходу і виходу, а також початковим станом автомата.

Під алфавітом у теорії автоматів розуміють непорожню множину різних символів.

Елементом алфавіту є букви.

Словом називається послідовність букв.

Абстрактний автомат представляється у вигляді чорного ящика із вхідними і вихідними алфавітами.

Математична модель цифрового автомата визначається компонентною функцією:

$$A = \{x, y, S, \delta, - \setminus, S_0\}$$

де x – вхідний алфавіт;
 y – алфавіт вихідних сигналів;
 S – множина станів автомата;
 δ – функція переходу;
 λ – функція виходу.

Під станом ми розуміємо опис або поведінку схеми, які залежать не тільки від входів, а й від попереднього стану.

4. Завдання автоматів. Таблиці і матриці переходів і виходів. Об'єднана таблиця

З точки зору опису цифрових автоматів використовуються різноманітні засоби, звані *мовами*.

Мови поділяються на **початкові та автоматні**.

Початкові мови дозволяють встановити логіку роботи цифрового автомата і послідовність здійснення подій.

Автоматні мови дозволяють встановити взаємозв'язок між станами автомата, вхідними і вихідними сигналами.

Способи опису цифрових автоматів поділяються на:

- Початкові мови:
 - ЛСА – логічна схема алгоритму;
 - ГСА – графічна схема алгоритму;
 - МСА – матрична схема алгоритму;
 - ФМП – функціональна мікропрограма;
 - СФП – система формул переходу.
- Автоматні мови:
 - ТП – таблиця переходів;
 - ТБ – таблиця виходів;
 - СТПіВ – спільна таблиця переходів і виходів;
 - МП – матриця переходів;
 - МВ – матриця виходів;
 - СМПіВ – спільна матриця переходів і виходів.

Автоматні мови мають три форми запису:

Таблична:

- форма представлена у вигляді таблиці, де стовпці позначаються станами автомата, а рядки – вхідними сигналами.

- На перетині рядків і стовпців встановлюються **стани** автоматів або вихідні сигнали.

Матрична:

- рядки і стовпці нумеруються станами, а на перетині рядків і стовпців записуються вхідні або вихідні сигнали.
- У рядку матриці вхідні і вихідні сигнали не повинні повторюватися двічі.

Графічна:

- послідовність роботи автомата виражається у вигляді графа, вершинами графа є стани автомата, перехід з одного стану в інший записується у вигляді дуги переходу, причому, **на початку** записується вхідний сигнал, а в кінці – вихідний.

Питання для самоконтролю

1. Що таке автомат, цифровий автомат?
2. Що таке детерміновані автомати?
3. Назвіть типи автоматів в залежності від особливостей вхідної стрічки і пам'яті.
4. Назвіть склад і опис автоматів.
5. Наведіть класифікацію автоматів за ознаками.
6. Що таке абстрактний автомат, алфавіт входу, виходу станів?
7. Назвіть способи опису цифрових автоматів.
8. Форми запису автоматних мов.

Тема 4. Теорія розпізнавання образів

План

1. Основні положення розпізнавання образів
2. Методи розпізнавання образів

1. Основні положення розпізнавання образів

Теорія розпізнавання образів (Pattern recognition) – розділ кібернетики, присвячений створенню теоретичних основ класифікації та ідентифікації предметів, явищ, процесів, сигналів, ситуацій тощо, які будемо узагальнено називати об'єктами, а процеси класифікації та ідентифікації – розпізнаванням [24, с.5].

Розпізнавання образів (pattern recognition) – це розділ теорії штучного інтелекту (artificial intelligence), що вивчає методи класифікації об'єктів. За традицією *об'єкт*, що піддається класифікації, називається *образом* (pattern). *Образом* може бути цифрова фотографія (розпізнавання зображень), буква або цифра (розпізнавання символів), запис мови (розпізнавання мови) тощо.

Розпізнавання образів (об'єктів, сигналів, ситуацій, явищ або процесів) – завдання ідентифікації об'єкта або визначення будь-яких його властивостей по його зображенню (оптичне розпізнавання) або аудіозапису (акустичне розпізнавання) та інші характеристики.

Образи володіють характерними об'єктивними властивостями в тому сенсі, що різні люди, що навчаються на різному матеріалі спостережень, здебільшого однаково і незалежно один від одного класифікують одні й ті самі об'єкти. У класичній постановці задачі розпізнавання універсальної маси розбивається на частини-образи.

Методика віднесення елемента до якого-небудь образу називається вирішальним правилом. Ще одне важливе поняття – метрика, спосіб визначення відстані між елементами універсальної множини. Чим менше ця відстань, тим більше схожими є об'єкти (символи, звуки та ін.) – те, що ми розпізнаємо. Зазвичай елементи задаються у вигляді набору чисел, а метрика – у вигляді функції. Від вибору представлення образів і реалізації метрики залежить ефективність програми, один алгоритм розпізнавання з різними метриками буде помилятися з різною частотою.

Приклади завдань розпізнавання образів:

- розпізнавання букв;

- розпізнавання штрих-кодів;
- розпізнавання автомобільних номерів;
- розпізнавання осіб та інших біометричних даних;
- розпізнавання мови [7, с. 26].

Прецедент – це об’єкт, приналежність якого до заданого класу визначена заздалегідь. Прецедентом може бути, наприклад, набір ознак пацієнта із відомим діагнозом, з яким слід порівнювати набір ознак людини, діагноз якої ще невідомий. Кожен образ являє собою набір чисел, що описують його властивості і називаються ознаками (feature).

Ознакою (Feature) розпізнавання називається характеристика певної властивості об’єкта, що аналізується.

Упорядкований набір ознак об’єкта називається вектором ознак (feature vector).

Вектор ознак – це точка в просторі ознак (feature space). Класифікатор, або вирішальне правило (decision rule) – це функція, яка ставить у відповідність вектору ознак образ класу, до якого він належить.

Задачу розпізнавання образів можна розділити на ряд підзадач.

1. Генерування ознак (feature generation) – вимірювання або обчислення числових ознак, що характеризують об’єкт.

2. Вибір ознак (feature selection) – визначення найбільш інформативних ознак для класифікації (в цей набір можуть входити не лише первинні ознаки, але й функції від них).

3. Побудова класифікатора (classifier construction) – конструювання вирішального правила, на підставі якого здійснюється класифікація.

4. Оцінка якості класифікації (classifier estimation) – обчислення показників правильності класифікації (точність, чутливість, специфічність, помилки першого та другого роду). Введемо позначення і сформулюємо математичну постановку задачі класифікації.

Задача класифікації *без учителя* (unsupervised classification) часто називається *кластеризацією* (clusterization), або кластерним аналізом (cluster analysis). В цій задачі вибірка образів, розбивається на підмножини, що не перетинаються (кластери), які складаються із схожих один на одного об’єктів, до того ж вимагається, щоб об’єкти із різних кластерів істотно відрізнялися один від одного.

Розглянемо приклад реальної задачі із області медичної діагностики. У 1962 р. Н. Nieburgs, а потім й інші дослідники, повідомили про характерні зміни в клітинах буккального епітелію пацієнта з пухлиною, локалізованою поза порожниною рота, і назвали це змінами, асоційованими з малігнізацією (МАС – malignancy associated changes). Ці зміни характеризуються зростанням розмірів ядер, розривами ядерної мембрани, зростанням розмірів зон, пов'язаних із хроматином, оточеним світлими областями. Ці автори зареєстрували схожі зміни при наявності передпухлинних процесів в організмі. У 2009 році пухлинно-асоційовані зміни були індуковані експериментально у мишей, яким привили злоякісні пухлини. Було висунуто гіпотезу, що імунна система реагує на появу злоякісних клітин, виробляючи фактор запалення (inflammation factor), який руйнує ДНК і провокує її згущення. Отже, зареєструвавши такі пухлинно-асоційовані зміни, ми могли б діагностувати рак.

2. Методи розпізнавання образів

У цілому, можна виділити три методи розпізнавання образів:

Метод перебору. У цьому випадку проводиться порівняння з базою даних, де для кожного виду об'єктів представлені всілякі модифікації відображення.

Наприклад, для оптичного розпізнавання образів можна застосувати метод перебору виду об'єкта під різними кутами, масштабами, зміщеннями, деформаціями і т. д. Для літер треба перебирати шрифт, властивості шрифту і т. д. У випадку розпізнавання звукових образів, відповідно, відбувається порівняння з деякими відомими шаблонами (наприклад, слово, вимовлене декількома людьми).

Другий підхід – проводиться більш глибокий аналіз характеристик образу. У разі оптичного розпізнавання це може бути визначення різних геометричних характеристик. Звуковий зразок у цьому випадку піддається частотному, амплітудному аналізу і т. д.

Наступний метод – використання штучних нейронних мереж (ШНМ). Цей метод вимагає або великої кількості прикладів задачі розпізнавання при навчанні, або спеціальної структури нейронної мережі, що враховує специфіку даної задачі. Тим не менш, його відрізняє більш висока ефективність і продуктивність [14].

Виділяють такі типи завдань розпізнавання:

- завдання розпізнавання – віднесення пред'явленого об'єкта з його опису до одного із заданих класів (навчання з учителем);
- завдання автоматичної класифікації – розбиття множини об'єктів, ситуацій, явищ за їх описами на систему непересічних класів (таксономія, кластерний аналіз, самонавчання);
- завдання вибору інформативного набору ознак при розпізнаванні;
- завдання приведення вихідних даних до виду, **зручного** для розпізнавання;
- динамічне розпізнавання і динамічна класифікація – задачі 1 і 2 для динамічних об'єктів;
- завдання прогнозування – суть попередній тип, в якому рішення має ставитися до деякого моменту в майбутньому.

В системах розпізнавання використовують ознаки:

- детерміновані;
- ймовірнісні;
- логічні; структурні;
- комбіновані.

Робочий словник ознак – це перелік ознак, які обрані з початкового словника ознак і використовуються у системі розпізнавання образів.

Клас образів – це підмножина образів з найбільш схожими суттєвими ознаками, обрана з множини однотипних образів за допомогою робочого словника ознак (однотипні образи: «Автомобілі» розділяються на класи «Волга», «Лада», «Запорожець» і т.д.).

Сукупність однотипних класів складають *алфавіт класів*.

Датчик – прилад, призначений для сприйняття, вимірювання, зберігання і переробки однієї або кількох ознак зовнішнього образу.

Датчики ознак образу: фізіологічний датчик: око людини, датчики переміщення і розмірів, датчики температури, датчики зображення і випромінювання, комп'ютерні візуальні датчики двомірних зображень, сканер, гама-томографія, ядерний магнітний резонанс, тактильні датчики, акустичні датчики, датчик Холла, електромагнітний датчик розпізнавання автомобілів.

Питання для самоконтролю

1. Що таке теорія розпізнавання образів?
2. Що таке образ?
3. Назвіть приклади завдань розпізнавання образів?
4. Наведіть методи розпізнавання образів?
5. Які ознаки використовують в системах розпізнавання?
6. Що таке робочий та початковий словник?

Тема 5. Кібернетика

План

1. Поняття кібернетики
2. Сфера застосування кібернетики
3. Основні напрямки кібернетики

1. Поняття кібернетики

Кібернетика (від др. -греч. κυβερνητική – «Мистецтво управління») – це наука про управління складними системами зі зворотним зв'язком.

Кібернетика – наука про процеси управління в складних динамічних системах, що ґрунтується на теоретичному фундаменті математики і логіки (взагалі на формальних мовах), а також на застосуванні обчислювальної техніки (комп'ютерів).

«Чорний ящик» це термін, який використовується у техніці й кібернетиці для позначення об'єкта чи системи, про принципи дії яких нічого невідомо, крім того, що певному вхідному сигналу відповідає певний вихідний сигнал [Вікіпедія].

Термін **«кібернетика»** в сучасному розумінні як наука про загальні закономірності процесів управління і передачі даних в машинах, живих організмах і суспільстві вперше був запропонований Норбертом Вінером в 1948 р.

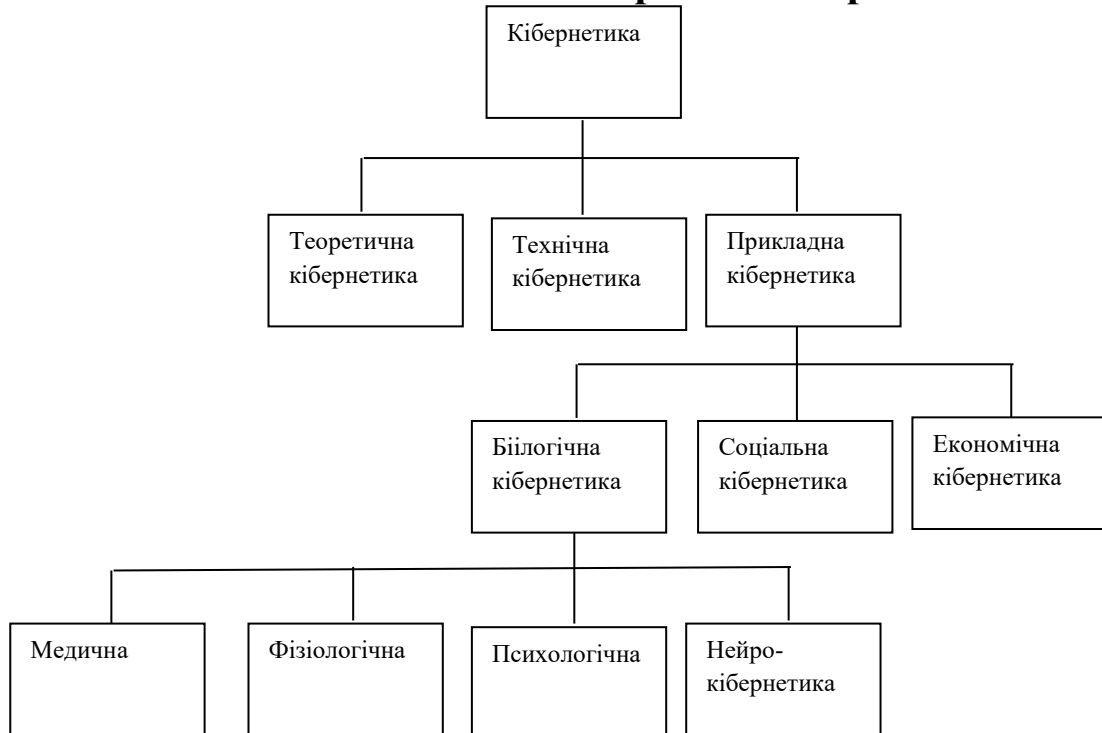
2. Сфера використання кібернетики

Кібернетика є міждисциплінарною наукою. Вона виникла на стику математики, логіки, семіотики, фізіології, біології, соціології. Їй притаманний аналіз та виявлення загальних принципів і підходів у процесі наукового пізнання.

Найбільш вагомими теоріями, що об'єднуються з кібернетикою, можна назвати наступні:

- теорія передачі сигналів;
- теорія управління;
- теорія автоматів;
- теорія прийняття рішень;
- теорія алгоритмів;
- розпізнавання образів;
- теорія оптимального управління.

3. Основні напрямки кібернетики



Теоретична кібернетика, подібно до математики є, власне кажучи, **абстрактною наукою**.

Завдання – розробка наукового апарату й методів дослідження систем управління незалежно від їхньої конкретної природи.

У теоретичну кібернетику увійшли й одержали подальший розвиток такі **розділи прикладної математики** як:

- *теорія інформації,*
- *теорія алгоритмів,*
- *теорія ігор,*
- *теорія масового обслуговування,*
- *дослідження операцій та ін.*

Технічна кібернетика — наука про управління технічними системами. Технічну кібернетику часто ототожнюють із сучасною теорією автоматичного регулювання й управління. Ця теорія, звичайно, служить важливою складовою частиною технічної кібернетики, але остання, разом з тим, включає питання розробки і конструювання автоматів (у тому числі сучасних ЕОМ і роботів), а також проблеми технічних засобів збирання, передачі, зберігання й перетворення інформації, розпізнання образів і т. д.

Прикладна кібернетика, у залежності від типу досліджуваних систем управління, поділяється на:

- *біологічну кібернетику,*
- *соціальну кібернетику,*

- *економічну кібернетику.*

Біологічна кібернетика вивчає загальні закони зберігання, передачі і переробки інформації у біологічних системах. Біологічну кібернетику у свою чергу підрозділяють на:

- медичну,
- фізіологічну,
- нейрокібернетику,
- психологічну.
- медична, котра займається, головним чином, моделюванням захворювань і використанням цих моделей для діагностики, прогнозування і лікування;
- фізіологічна, що вивчає і моделює функції клітин і органів у нормі та патології;
- нейрокібернетика, в якій моделюються процеси переробки інформації в нервовій системі;
- психологічна, що моделює психіку на основі вивчення поведінки людини.

Проміжною ланкою між **біологічною** та **технічною кібернетикою** є **біоніка** - наука про використання моделей біологічних процесів і механізмів як прототипів для удосконалювання існуючих і створення нових технічних пристроїв.

Соціальна кібернетика – наука, у якій використовуються методи і засоби кібернетики з метою дослідження й організації процесів управління в соціальних системах.

Економічна кібернетика, сферою якої є проблеми оптимізації управління економікою в цілому, її окремими галузями, економічними районами, промисловими комплексами, підприємствами і т.д.

Основне **прикладне призначення кібернетики** – проектування автоматичних, автоматизованих та інтегрованих систем різного класу і призначення.

Рівні предметної галузі кібернетики:

- загальнодержавна автоматизована система збору та обробки інформації (ЗДАС);
- територіальні (республіканські, обласні, міські, районні) автоматизовані системи обробки інформації і управління (ТАСУ);
- галузеві автоматизовані системи управління (ГАСУ);
- автоматизовані системи управління акціонерними товариствами, підприємством (АСУП);

- ☐ автоматизовані установчі системи (АУС);
- ☐ автоматизовані робочі місця керівників (АРМ).

Одним з основних методів кібернетики є метод математичного моделювання систем і процесів управління. Сенс його полягає в тому, що експерименти проводяться не з реальною фізичною моделлю об'єкта, що вивчається, а з його математичним описом, реалізованим в комп'ютері.

Метою кібернетики є оптимізація систем управління.

Питання для самоконтролю

1. Що таке кібернетика?
2. Назвіть сфери використання кібернетики.
3. Наведіть напрями кібернетики.
4. Хто запропонував термін кібернетика?

ЗМІСТОВИЙ МОДУЛЬ II. САМОСТІЙНА РОБОТА

Вимоги до самостійної роботи

1. Самостійна робота складається з 8 завдань.
2. Всі розв'язки повинні містити детальний опис ходу розв'язування завдання з наведенням умови, всіх використаних формул та остаточної відповіді. Для формування умови кожного завдання необхідно в загальні умови самостійної роботи підставити власні значення параметрів k (номер студента у списку групи) та l (номер дня в межах місяця з дати народження студента) та заокруглити числа до більшого (три цифри після коми). Всі логарифми у самостійній роботі беруться з основою 2.
3. Розв'язки завдань самостійної роботи здаються у рукописному вигляді на листках формату А4. Титульна сторінка самостійної роботи оформляється за зразком, наведеним наприкінці цих завдань та здається у друкованому вигляді. Листки самостійної роботи не скріплюються, а вкладаються в папку-конверт.
4. Під час захисту самостійної роботи студент має продемонструвати розуміння матеріалу та підтвердити самостійність її виконання.
5. Наприкінці роботи наводиться список використаних джерел.

Теми самостійної роботи

№ з/п	Назва теми
1.	Теорія інформації
2.	Теорія кодування
3.	Теорія автоматів
4.	Теорія розпізнавання образів
5.	Математична кібернетика

Варіант №1

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Вар.	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
1	0,15	0,07	0,09	0,08	0,09	0,12	0,04	0,13	0,12	0,10

2. Закодувати повідомлення МОРЕПРОДУКТИ алгоритмом Шеннона-Фано шляхом побудови графа.

3. Записати десяткове число 221 у двійковій системі числення, та 443 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Яка кількість вільного місця має бути у електронній поштовій скриньці, щоб можна було отримати 3 повідомлення.

Кожне таке повідомлення містить по 3500 символів (1 символ займає 8 біт). Ще кожне таке повідомлення містить в собі також по 2 прикріплених файла з графічною інформацією розміром по 75 Кбайт кожний.

6. Закодуйте/декодуйте фрази *We know the truth* шифром Цезаря.

7. Закодуйте/декодуйте фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **The human soul tight and taiga** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 2

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
2	0,11	0,08	0,13	0,10	0,08	0,08	0,13	0,10	0,09	0,11

2. Закодувати повідомлення **ПРОМИСЛОВІСТЬ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 875, та 413 у вісімковій системі числення з перевітками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Яку мінімальну пропускну здатність (мегабіт за секунду) повинна мати локальна комп'ютерна мережа для забезпечення передачі за три секунди текстової документації об'ємом в 190 сторінок. Відомо, що одна сторінка документації містить 29 рядків по 67 символів у рядку (1 символ займає 8 біт).

6. Закодувати/декодувати фрази *It's a good beast, without it, life is bad* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **But in the heart of summer** – а вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 3

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з табл. згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11

3	0,09	0,11	0,09	0,11	0,13	0,08	0,11	0,08	0,09	0,09
---	------	------	------	------	------	------	------	------	------	------

2. Закодувати повідомлення **ПІДВИЩЕННЯ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 345, та 143 у вісімковій системі числення з перевітками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею тим методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Відомості про кожного працівника підприємства зберігаються в електронній базі даних відділу кадрів у вигляді тексту в середньому по 4048 символів (1 символ займає 8 біт). На якій мінімальній кількості дискет ємністю 1,4 Мбайт можна

розмістити відомості про всіх 1456 працівників цього підприємства.

6. Закодувати/декодувати фрази *More in the land of poets* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **A wise and triumphant genius kind** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 4

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
4	0,10	0,09	0,06	0,07	0,12	0,15	0,07	0,09	0,13	0,10

2. Закодувати повідомлення **СИСТЕМАТИЧНИЙ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 678, та 321 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. З якою регулярністю (днів) необхідно видаляти повідомлення із електронної поштової скриньки, яка має об'єм 2 Мбайт, якщо кожний день на неї надходить 15 повідомлень. Відомо, що кожне таке повідомлення містить по 1500 символів (1 символ займає 8 біт). Також, кожне таке повідомлення містить по два прикріплених файли по 10 Кбайт кожний.

6. Закодувати/декодувати фрази *But it is not necessary to call him* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **And this performance, he is no** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 5

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з табл. згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
5	0,14	0,08	0,09	0,05	0,08	0,14	0,12	0,09	0,12	0,09

2. Закодувати повідомлення **ІНТЕРКОНТИНЕНТАЛЬНИЙ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 967, та 321 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Швидкість вводу текстової інформації із клавіатури оператором ПК з екрана монітора становить 70 слів за хвилину (одне слово містить в середньому шість символів (1 символ займає 8 біт). Який об'єм тексту (кілобайт) встигне ввести оператор ПК за три години безперервної роботи?

6. Закодувати/декодувати фрази *Another moment, only this moment* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **A passionate dream lover summer** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 6

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
6	0,13	0,06	0,15	0,12	0,13	0,09	0,07	0,07	0,05	0,12

2. Закодувати повідомлення **ДЕРЖГІРПРОМНАГЛЯД** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 845, та 193 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею тим методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів

для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Який розмір повинна мати електронна поштова скринька, щоб на неї могло надходити без очищення 20 електронних повідомлень. Кожне таке повідомлення містить в собі 4503 символи (1 символ займає 8 біт) і один прикріплений файл об'ємом 60 Кбайт.

6. Закодувати/декодувати фрази *He will come himself and will not* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **He traps waiting for sight words** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 7

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
7	0,06	0,13	0,08	0,13	0,10	0,07	0,10	0,14	0,09	0,09

2. Закодувати повідомлення **ВИРОБНИЦТВО** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 278, та 643 у вісімковій системі числення з перевітками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею тим методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Середня швидкість читання текстового документа оператором ПК з екрана монітора становить 120 слів за хвилину (одне слово в середньому — 5 символів (1 символ займає 8 біт)). Скільки часу потрібно оператору ПК для прочитування електронного повідомлення, що надійшло по електронній пошті і має об'єм 28 Кбайт.

6. Закодувати/декодувати фрази *All of the whole forest* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **He hears the quiet steps** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 8

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
8	0,10	0,10	0,08	0,10	0,07	0,09	0,12	0,11	0,13	0,10

2. Закодувати повідомлення **ВИРОБНИЦТВО** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 175, та 140 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Яку максимальну кількість сторінок текстової документації можна помістити на оптичний диск розміром 5,2 дюйма з метою створення резервної копії. Текст на одній сторінці розміщується у 39 рядках по 64 символи в кожному (1 символ займає 8 біт). Оптичний диск (CD-ROM) вміщує 700 Мбайт інформації.

6. Закодувати/декодувати фрази *Evening dream lover summer* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **It's still summer sleeps, and the** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 9

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
9	0,12	0,08	0,11	0,11	0,09	0,14	0,14	0,09	0,07	0,05

2. Закодувати повідомлення **МОЖНОВЛАДЦІ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 112, та 984 у вісімковій системі числення з перевітками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницеvim методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. Опис видів сільськогосподарської продукції та послуг підприємства розміщується на 130 сторінках (кожна сторінка тесту має 21 рядок по 64 символи в кожному (1 символ займає 8 біт). Скільки часу необхідно для передачі такого документа по електронній мережі, пропускна здатність якої становить 1 Мбіт/с.

6. Закодувати/декодуйте фрази *I'm a tree, I am snow, I am all that* шифром Цезаря.

7. Закодуйте/декодуйте фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **In these gardens, in the sonatas** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Варіант 10

1. Знайти ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці згідно варіанту.

Варі- ант	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
10	0,11	0,08	0,08	0,14	0,08	0,11	0,08	0,07	0,14	0,13

2. Закодувати повідомлення **ПЕРЕТВОРЕННЯ** алгоритмом Шеннона-Фано шляхом побудови графа. Завдання згідно варіанту.

3. Записати десяткове число у двійковій системі числення 429, та 583 у вісімковій системі числення з перевірками.

4. А) Розпізнавання образів. Перетворити на коди перші 4 літери власного прізвища студента для рецепторного поля у 6 рядків та 6 колонок. Визначити відстані між ними. Розділити на 2 класи в залежності від відстані до 1-ї літери.

Б) Перетворити у код власний чорно-білий малюнок у рецепторному полі з 36 комірок. Ввести у 5 комірок «перешкоди». Порівняти інформацію з внесеними «перешкодами» та без внесених «перешкод».

В) Задайте 7 параметрів ознак, які контролюються: $?_1 = ? ; ?_2 = ? ; ?_3 = ? ; ?_4 = ? ; ?_5 = ? ; ?_6 = ? ; ?_7 = ? ;$. Різницею методом кодування сигналів отримати код.

Г) Виконати умовні контури двох різних за розмірами (формою) літаків на площі (X, Y) . Методом центральних моментів для кодування ознак рухомих об'єктів отримати чотири ознаки для кожного літака.

Д) Скласти довільні списки X_e та X_k , кожний з яких має не менше 6-ти цифрових компонентів. Отримати коди і розрахувати міру близькості між цими двома об'єктами методом кодування та визначення міри близькості між об'єктами з ознаками «1» та «0».

Е) Описати три початкові літери прізвища студента не менш ніж 4-ма суттєвими якісними ознаками, що приймають значення

або «1», або «0». Приклад якісних ознак: «Чи є горизонтальна риска внизу/вверху літери?», «Чи є вертикальна риска справа/зліва?», «Чи є верхня/нижня риска під кутом $\pm 45^\circ$?». Скласти вимоги до датчиків. Розділити три літери на два класи.

5. З якою мінімальною продуктивністю (Кбіт за секунду) має працювати модем, для забезпечення завантаження однієї сторінки з мережі Internet за 10 секунд. Відомо, що ця сторінка містить 4000 символів (1 символ займає 8 біт) та 3 графічних об'єкта (рисунок) по 9 Кбайт кожний.

6. Закодувати/декодувати фрази *Beautiful woman, tenderness warm* шифром Цезаря.

7. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

8. Зашифруйте фразу **Don't put a hand on his shoulder** вибравши її самостійно методом шифру підстановок (квадрат Полібія).

Модульна контрольна робота
з дисципліни «Теоретичні основи інформатики»

ПІБ студента _____
Група / курс _____
Дата проведення _____

ВАРІАНТ № 1

№ з/п	Питання, завдання	К-ть балів за завдання	Набрана к-ть балів
<i>Питання I рівня складності</i>			
1.	Кібернетика це - А) це наука про управління складними системами зі зворотним зв'язком; Б) розвиває теоретичні основи й методи класифікації та ідентифікації предметів, явищ, процесів, сигналів, ситуацій і т. д. об'єктів, які характеризуються кінцевим набором деяких властивостей і ознак; В) розділ кібернетики, де за допомогою математичних методів вивчаються способи вимірювання кількості інформації, що містяться в будь-якому повідомленні; Г) сукупність технічних засобів, що забезпечують передачу повідомлення від джерела до одержувача по одній лінії зв'язку.	16.	
2.	Властивості інформації : А) об'єктивність; Б) корисність; В) логічність; Г) зрозумілість; Д) економічність; Е) повнота; Є) стабільність.	0,5 б.	
3.	Сучасні інформаційні системи поділяються на: А) неавтоматизовані, автоматизовані, безперервні; Б) дискретні, безперервні, квантові; В) неавтоматизовані, автоматизовані, автоматичні;	0,5 б.	

	Г) ручні, автоматизовані, автоматичні.		
4.	Розрізняють сигнали: А) акустичні, електричні, оптичні, гідравлічні; Б) акустичні, електричні, ігрові, гідравлічні; В) акустичні, електричні, ігрові, гідравлічні; Г) статичні, електричні, ігрові, гідравлічні.	16.	
5.	Каналом зв'язку називається -... А) сукупність технічних засобів, що забезпечують передачу повідомлення від джерела до одержувача по одній лінії зв'язку; Б) сукупність зв'язків, що забезпечують передачу повідомлення від одержувача до роздавача; В) ланцюжок сигналів визначеної довжини, що дозволяють створити з'єднання даних; Г) ланцюжок сигналів визначеної довжини, що позначають певну інформацію.	16.	
6.	Кодування – це... А) сукупність технічних засобів, що забезпечують передачу повідомлення від джерела до одержувача по одній лінії зв'язку; Б) процес присвоювання інформації умовних позначень за допомогою певного правила; В) як математична модель, описує реальні технічні автомати пристроїв; Г) ланцюжок сигналів визначеної довжини, що дозволяють створити з'єднання даних.	16.	
7.	Сигнал - це ... А) зміна стану речовини чи поля, що передається на певну віддаль; Б) сукупність технічних засобів, що забезпечують передачу повідомлення від джерела до одержувача по одній лінії зв'язку; В) сукупність зв'язків, що забезпечують передачу повідомлення від одержувача до роздавача; Г) ланцюжок сигналів визначеної довжини, що дозволяють створити з'єднання даних.	16.	
8.	Алгоритм префіксного неоднорідного кодування це – А) Алгоритм Лемпела-Зіва-Велча; Б) кодування Шеннона-Фано; В) коду Хаффмана;	16.	

	Г) штрихові коди.		
9.	Теорія автоматів – це А) сукупність технічних засобів, що забезпечують передачу повідомлення від джерела до одержувача по одній лінії зв'язку; Б) наука про процеси управління в складних динамічних системах, що ґрунтується на теоретичному фундаменті математики і логіки; В) аналогове повідомлення, представляється фізичною величиною, що передає безперервне повідомлення; Г) пристрої, призначені для обробки і перетворення цифрової інформації з метою управління.	16.	
10.	За обсягом пам'яті цифрові автомати бувають: А) з пам'яттю - послідовні і без пам'яті; Б) Милі та Мура; В) абстрактні та структурні; Г) підавтомат і надавтомат.	16.	
11.	За кінцівками множин входних і вихідних сигналів, і множинами станів, цифрові автомати бувають: А) з пам'яттю - послідовні і без пам'яті; Б) Милі та Мура; В) кінцеві автомати і нескінченні автомати; Г) підавтомат і надавтомат.	16.	
12.	Абстрактний автомат - це модель цифрового автомата ... А) яка характеризується алфавітом входу, алфавітом виходу, алфавітом станів, функціями переходу і виходу, а також початковим станом автомата; Б) в якому керуючий автомат управляє блоками об'єктів управління, грубо кажучи, об'єкт управління не один; В) для якого керуючий автомат реалізує функції управління всіма блоками, тобто блок не єдиний; Г) призначений для обробки і перетворення цифрової інформації з метою управління.	16.	
<i>Питання II рівня складності (короткий опис відповіді)</i>			
13.	Назвіть властивості ентропії	2 б.	
14.	Назвіть та опишіть форми запису автоматних мов	2 б.	
<i>Питання III рівня складності (здійснити розв'язання завдань)</i>			

15	Який розмір повинна мати електронна поштова скринька, щоб на неї могло надходити без очищення 20 електронних повідомлень. Кожне таке повідомлення містить в собі 4503 символи (1 символ займає 8 біт) і один прикріплений файл об'ємом 60 Кбайт.	5 б.	
----	--	------	--

Всього набрано балів:

Оцінка _____

Підпис викладача _____

Екзаменаційні питання

1. Інформатика. ЕОМ. Покоління ЕОМ.
2. Форми подання, носії інформації. Типи інтерфейсів.
3. Характеристика розділів інформатики.
4. Концепції теорії інформації. Одиниці вимірювання інформації.
5. Принципи отримання, зберігання, обробки та використання інформації.
6. Ентропія. Властивості інформації. Умовна ентропія, властивості умовної ентропії. Безумовна ентропія.
7. Інформація. Властивості інформації. Сигнал. Канал передачі інформації.
8. Інформаційні процеси. Інформаційні системи. Різновиди сучасних інформаційних систем.
9. Кодування. Код. Декодування. Приклади кодування.
10. Кількісні характеристики інформації. Носії інформації.
11. Системи числення. Характеристика кодів. Спосіб подання кодів.
12. Кодування Шенонна-Фано.
13. Шифрування. Дешифрування. Штрихові коди.
14. Приклади шифрування інформації.
15. Стиснення даних. Способи кодування інформації.
16. Алгоритм Лемпела-Зіва-Велча.
17. Автомати. Теорія автоматів. Кінцеві автомати.
18. Цифрові (дискретні) автомати. Види інформації з якими працює автомат.
19. Особливості цифрових обчислювальних машин (автоматів).
20. Детерміновані автомати. Характеристика детермінованих автоматів.
21. Типи автоматів у залежності від особливостей вхідної стрічки і пам'яті.
22. Модель автомата; склад і опис.
23. Абстрактний автомат, способи призначення об'єктів.
24. Аспекти роботи автоматів.
25. Цифрові автомати. Класифікація цифрових автоматів.
26. Мікропроцесорні цифрові автомати і структури їх побудови.
27. Етапи процесу розробки автоматів.

28. Структурні автомати. Структура побудови багатоблокових автоматів.
29. Визначення абстрактного автомата. Алфавіти входу, виходу, станів. Функції виходів і переходів.
30. Завдання автоматів. Таблиці і матриці переходів і виходів. Об'єднана таблиця. Графи автоматів.
31. Способи опису цифрових автоматів.
32. Форми запису автоматних мов.
33. Автомат Мура. Закон функціонування автомата Мура.
34. Автомат Мілі. Закон функціонування автомата Мілі.
35. Мінімізація автоматів. Ступінь сумісності стану.
36. Етапи заповнення трикутної матриці.
37. Основні поняття розпізнавання образів.
38. Системи оптичного розпізнавання. Алфавіт класів.
39. Первинний словник ознак. Робочий словник ознак.
40. Системи, які вирішують задачу розпізнавання.
41. Методи класифікація розпізнавання образів.
42. Історія теорії розпізнавання образів.
43. Образи. Приклади образів. Прецедент. Ознака. Вектор ознак.
44. Характеристика задач розпізнавання образів.
45. Опис ознак в системах розпізнавання образів.
46. Робочий словник ознак. Клас образів.
47. Засоби, які використовуються при розпізнаванні образів.
48. Структура системи розпізнавання образів.
49. Класифікація методів розпізнавання образів.
50. Поняття кібернетики. Чорний ящик.
51. Сфера застосування кібернетики.
52. Основні напрямки кібернетики.
53. Енерджентність. Провідні фахівці кібернетики.
54. Прикладна кібернетика. Характеристика прикладної кібернетики.
55. Біологічна кібернетика. Характеристика біологічної кібернетики.
56. Чиста кібернетика. Рівні предметної області кібернетики:
57. Задачі. Експериментальне визначення. Ентропії повідомлення.
58. Джерела А та В мають розподіли $P_A = \{0,1; 0,1; 0,15; 0,125; 0,125; 0,1; 0,15; 0,15\}$, $P_B = \{0,5; 0,3; 0,1; 0,025; 0,025; 0,02; 0,15; 0,15\}$. Ентропія якого джерела більше?

59. Класифікація систем розпізнавання образів.

60. Для об'єкта (принтери) визначити алфавіт класів (2-3 класи) та початковий і робочий словники ознак класів. Визначити датчики, які потрібно використовувати для визначення ознак об'єктів. Вказати вимоги до нормалізації та квантування сигналів.

61. Записати десяткове число 333 у двійковій системі числення.

62. Закодувати/декодувати фрази шифром Цезаря фразу «More in the land of poets».

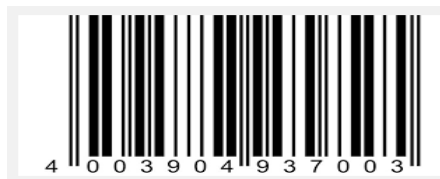
63. Закодувати/декодувати фразу шифром Віженера. Ключове слово і кодовану фразу придумати самостійно (латиницею).

64. Зашифрувати/дешифрувати фразу (латиницею), вибравши її самостійно методом шифру підстановок (квадрат Полібія).

65. Розглянути штрих-код 4003904937003 і визначити:

1) справжність товарів, розрахувавши контрольний знак штрих-коду;

2) чи всі вимоги до зображення ШК на цих товарах дотримані?



66. Закодувати/декодувати фразу «All of the whole forest» шифром Цезаря.

67. Який розмір повинна мати електронна поштова скринька, щоб на неї могло надходити без очищення 20 електронних повідомлень. Кожне таке повідомлення містить в собі 4503 символи (1 символ займає 8 біт) і один прикріплений файл об'ємом 60 Кбайт.

68. Зашифрувати та дешифрувати фразу «Cogito, ergo sum» за допомогою квадрата Полібія.

ГЛОСАРІЙ

Brute force («грубою силою») – перебір всіх можливих варіантів (ключів).

Known plaintext (з відомим відкритим текстом) – ключ обчислюється за відомим фрагментом відкритого тексту та відповідного йому зашифрованого.

Автентичність даних і систем (authenticity of information) – для даних автентичність можна визначити як факт підтвердження дійсності інформації, що міститься в цих даних, а для систем – здатність забезпечувати процедуру відповідної перевірки – аутентифікації даних.

Аутентифікація (authentication) – процедура перевірки дійсності даних, пересвідчення чи доведення факту створення та відправлення легітимними (законними) учасниками процесу обміну інформації.

Блочні (блокові) шифри – шифрують інформацію строго визначеними порціями – блоками, наприклад, по 64 біти.

Гама-послідовність чи просто гама (gamma sequence, gamma) – послідовність псевдовипадкових елементів, що генеруються за певним законом чи алгоритмом.

Гамування (gamma xoring) – процес «накладання» гама-послідовності на відкриті дані, найчастіше з допомогою логічної операції XOR (виключне АБО).

Дешифрування (deciphering) і розшифровування (decryption) – відповідно методи одержання із зашифрованого тексту інформації без знання криптографічного ключа і зі знанням його.

Імітовставка – блок додаткової інформації, обчислений за певним алгоритмом і залежний від деякого криптографічного ключа і даних.

Імітозахист – це захист даних у системах передачі і зберігання від спотворення чи зловмисного нав'язування помилкової інформації.

Криптоаналіз (cryptanalysis) – набір методик і алгоритмів дешифрування криптографічно захищених повідомлень, аналізу шифросистем.

Криптоаналітична атака (cryptoanalytic attack) – спроба викликати відхилення від нормального проведення процесу конфіденційного обміну інформацією.

Криптографічна стійкість, криптостійкість (cryptographic strength) – стійкість шифросистеми проти всіх відомих видів криптоаналізу.

Криптографічний ключ (cryptographic key, іноді просто key) – для симетричних криптосистем – секретний компонент шифру.

Підключ шифрування (round key, subkey) – криптографічний ключ, що обчислюється і використовується тільки на етапі шифрування з основного ключа шифрування.

Потокові (поточні) шифри – шифрують інформацію по мірі надходження (побайтово чи побітово).

Принцип Керкхоффа (Kerckhoff) – принцип розробки і поширення криптографічних алгоритмів, відповідно до якого в секреті тримається тільки визначений набір параметрів шифру (і в обов'язковому порядку криптографічний ключ), а все інше може бути відкритим без зниження криптостійкості алгоритму.

Раунд чи цикл шифрування (round) – один комплексний крок алгоритму, в процесі якого перетворюються дані.

Симетрична криптосистема – система із секретним ключем, який використовується і під час шифрування, і під час розшифровування.

Статистичний – обчислення ключа ведеться із врахуванням статистичних закономірностей відкритого тексту (частота появи букв, цифр, складів).

Стеганографія – малопомітне приховування важливої інформації, переважно текстової, у великих за об'ємах графічних, звукових або текстових файлах, які передаються відкритими каналами.

Шифр і шифросистема (cipher, cypher, ciphercode) – найчастіше вихід криптосистеми і сама симетрична криптосистема відповідно.

Шифрування (encrytion) – процес зашифровування інформації, тобто застосування криптографічного перетворення до вхідних даних (відкритого тексту).

Додаток А

Титульна сторінка

Національний університет «Чернігівський колегіум»
імені Т.Г. Шевченка
кафедра інформатики і обчислювальної техніки

**Самостійна робота з дисципліни
Теоретичні основи інформатики**

Варіант ____

Виконав:

студент V курсу
спеціальності «Математика»
групи *шифр групи*

Прізвище, ім'я, по батькові студента

Перевірила:

доц. Матвійчук Л.А.

Чернігів - 2018

Перетворення десяткового числа (див. таблиця варіантів) у
двійкову систему числення

Варіант	Десяткове число	Варіант	Десяткове число	Варіант	Десяткове число
1	552, 743	8	121, 120	15,	134, 220
2	133, 134	9	231, 120	16	147, 230
3	650, 724	10	334, 125	17	358, 116
4	165, 170	11	342, 360	18	317, 560
5	130, 171	12	373, 170	19	132, 147
6	170, 264	13	321, 235	20	431, 126
7	165, 217	14	303, 423	21	151, 360
				22	123, 540

Перетворення десяткового число (див. таблиця варіантів) у
вісімкову систему числення

Варіант	Десяткове число	Варіант	Десяткове число	Варіант	Десяткове число
1	382, 443	8	120, 119	15,	234, 229
2	333, 234	9	230, 118	16	347, 228
3	150, 124	10	334, 115	17	348, 116
4	155, 180	11	345, 378	18	357, 280
5	170, 141	12	323, 470	19	432, 287
6	170, 224	13	331, 235	20	431, 286
7	183, 211	14	343, 228	21	111, 360
				22	123, 240

Додаток В.

Визначити кодову відстань між такими комбінаціями рівномірного трійкового коду завдовжки $n=3$, якщо значення А, В, С, D – взяти із таблиці

варіант	А, В, С, D значення	варіант	А, В, С, D значення
1	A = 012, B = 033, C = 210, D = 120.	12	A = 021, B = 212, C = 130, D = 010.
2	A = 020, B = 002, C = 002, D = 221.	13	A = 002, B = 212, C = 120, D = 010.
3	A = 024, B = 032, C = 210, D = 221.	14	A = 005, B = 001, C = 112, D = 130.
4	A = 031, B = 003, C = 123, D = 210.	15	A = 002, B = 003, C = 220, D = 110.
5	A = 210, B = 022, C = 111, D = 210.	16	A = 003, B = 030, C = 220, D = 201.
6	A = 001, B = 003, C = 120, D = 210.	17	A = 010, B = 020, C = 140, D = 230.
7	A = 050, B = 020, C = 120, D = 210.	18	A = 123, B = 030, C = 140, D = 201.
8	A = 002, B = 010, C = 145, D = 010.	19	A = 040, B = 010, C = 110, D = 240.
9	A = 201, B = 010, C = 020, D = 010.	20	A = 002, B = 011, C = 222, D = 121.
10	A = 001, B = 020, C = 222, D = 010.	21	A = 010, B = 020, C = 130, D = 121.
11	A = 003, B = 222, C = 020, D = 010.	22	A = 031, B = 123, C = 303, D = 111.

Закодувати/декодуйте фрази шифром Цезаря

<i>Var-m</i>	фраза для шифрування	<i>Var-m</i>	фраза для шифрування
1.	<i>We know the truth</i>	2.	<i>It's a good beast, without it, life is bad.</i>
3.	<i>More in the land of poets</i>	4.	<i>But it is not necessary to call him.</i>
5.	<i>Another moment, only this moment</i>	6.	<i>He will come himself and will not</i>
7.	<i>All of the whole forest</i>	8.	<i>Evening dream lover summer</i>
9.	<i>I'm a tree, I am snow, I am all that</i>	10.	<i>Beautiful woman, tenderness warm</i>
11.	<i>Don't put a hand on his shoulder.</i>	12.	<i>Ripe passion for fire and fetus</i>
13.	<i>In these gardens, in the sonatas</i>	14.	<i>The fire radiates melonoma on stage</i>
15.	<i>He hears the quiet steps</i>	16.	<i>It's still summer sleeps, and the</i>
17.	<i>He traps waiting for sight words</i>	18.	<i>a passionate dream lover summer</i>
19.	<i>And this performance, he is no</i>	20.	<i>A wise and triumphant genius kind</i>
21.	<i>The human soul tight and taiga</i>	22.	<i>But in the heart of summer – a</i>

Додаток Д.

Розглянути штрих-коди і визначити:

1) справжність товарів, розрахувавши контрольний знак штрих-кодів;

2) чи всі вимоги до зображення ШК на цих товарах дотримані?

<i>Вар-т</i>	Приклад	<i>Вар-т</i>	Приклад
1		12	
2		13	
3		14	
4		15	
5		16	
6		17	
7		18	

8		19	
9		20	
10		21	
11		22	

Кодування інформації у ШК EAN

Сим- воли	Набір А	Набір В	Набір С	Набір D
	Двійковий еквівалент	Двійковий еквівалент	Двійковий еквівалент	Двійковий еквівалент
0	0001101	0100111	1110010	1011000
1	0011001	0110011	1100110	1001100
2	0010011	0011011	1101100	1011110
3	0111101	0100001	1000010	1011110
4	0100011	0011101	1011100	1100010
5	0110001	0111001	1001110	1000110
6	0101111	0000101	1010000	1111010
7	0111011	0010001	1000100	1101110
8	0110111	0001001	1001000	1110110
9	0001011	0010111	1110100	1101000
H_1	101	Обмежувальні знаки СТАРТ і СТОП		
H_2	010101			
H_3	101010			
H_4	01010	Розділові знаки		
H_5	10101			

Набір літер яким кодується знаки лівої частини кодового слова

u_{12}	u_{11}	u_{10}	u_9	u_8	u_7	u_6
0	A	A	A	A	A	A
1	A	A	B	A	B	B
2	A	A	B	B	A	B
3	A	A	B	B	B	A
4	A	B	A	A	B	B
5	A	B	B	A	A	B
6	A	B	B	B	A	A
7	A	B	A	B	A	A
8	A	B	A	B	B	A
9	A	B	B	A	B	A

Додаток 3.

Двійкові логарифми цілих чисел

x	$\log_2 x$	x	$\log_2 x$	x	$\log_2 x$	x	$\log_2 x$
1	0,000	21	4,392	41	5,358	61	5,931
2	1,000	22	4,459	42	5,392	62	5,951
3	1,585	23	4,524	43	5,426	63	5,977
4	2,000	24	4,585	44	5,459	64	6,000
5	2,322	25	4,644	45	5,492	65	6,022
6	2,585	26	4,700	46	5,524	66	6,044
7	2,807	27	4,755	47	5,555	67	6,066
8	3,000	28	4,807	48	5,585	68	6,087
9	3,170	29	4,858	49	5,616	69	6,109
10	3,332	30	4,907	50	5,644	70	6,129
11	3,459	31	4,954	51	5,672	71	6,150
12	3,585	32	5,000	52	5,700	72	6,170
13	3,700	33	5,044	53	5,728	73	6,190
14	3,807	34	5,087	54	5,755	74	6,209
15	3,907	35	5,129	55	5,781	75	6,119
16	4,000	36	5,170	56	5,807	76	6,248
17	4,087	37	5,209	57	5,833	77	6,267
18	4,170	38	5,248	58	5,858	78	6,285
19	4,248	39	5,285	59	5,883	79	6,304
20	4,322	40	5,322	60	5,907	80	6,322

Список використаних джерел

1. Айзерман М. А. Метод потенційних функцій в теорії навчання машин. / М. А. Айзерман, Е. М. Браверман, Л. І. Розоноер. – М.: [Наука](#), 2004. – 384 с.
2. Блейхут Р. Теория и практика кодов, контролирующих ошибки / Р.Блейхут. М.: Мир, 1986. – 576 с.
3. Горбань А. Нейронні мережі на персональному комп'ютері / А. Горбань, Д. Россиев // [Новосибірськ](#), Наука, 1996. – С. 114–119.
4. Игнатов В. А. Теория информации и передачи сигналов / В. А. Игнатов. – М.: Радио и связь, 1992. – 280 с.
5. Журавльов Ю. І. Вибрані наукові праці / Ю. І. Журавльов. – Вид. Магістр, 2002. – 420 с.
6. Журавльов Ю. І. Про алгебраїчному підході до вирішення задач розпізнавання та класифікації / І.Ю. Журавльов // Проблеми кібернетики. М.: Наука, 2005. – Вип. 33. С. 5–68.
7. Жураковський Ю. П. Теорія інформації та кодування / Ю. П. Жураковський, В. П. Полторак. – К.: Вища шк., 2001. – 255с.
8. Закон України «Про вищу освіту» [Електронний ресурс] // Відомості Верховної Ради України. – 2002. – №20. – Режим доступу до журн.: <http://www.sumdu.edu.ua/ua/general/low/zakon/ZU2984.doc>
9. Електронна енциклопедія Wikipedia (Електронний ресурс) / Спосіб доступу: URL: <http://www.wikipedia.org/>. – Загол. з екрана.
10. Конспект лекцій до курсу «Теорія масової інформації» для студентів I курсу денної та заочної форми навчання спеціальності 6.030301 «Журналістика» / Укл.: О.В.Юферева. – Запоріжжя: ЗНТУ, 2015. – 76 с.
11. Кузьмин И. В. Основы теории информации и кодирования / И. В. Кузьмин, В. А. Кедрус, – К. Вища шк. Главное изд-во, 1986, – 238 с.
12. Ларина Н. А. Защита информации. Крптология: Методическое пособие для бакалавров направления подготовки 09.03.01 «Информатика и вычислительная техника» / Н. А. Ларина // Рубцовский индустриальный институт. – Рубцовск, 2014. – 56 с.

13. Мазуров В.Д. Комітети систем [нерівностей](#) та завдання розпізнавання / В.Д. Мазуров // [Кібернетика](#), 2004, № 2. С. 140-146.
14. Орлов В. А. Теория информации в упражнениях и задачах / В. А. Орлов, Л. И. Филиппов. – Москва, «Высшая школа», 1976, с. 230.
15. Потапов А. С. Розпізнавання образів та машинне сприйняття / Потапов А. С. – С-Пб.: Політехніка, 2007. – 548.
16. Растрігін Л. А. Метод [колективного](#) розпізнавання. / Л. А. Растрігін, Р. Х. Еренштейн. – М. Энергоіздат, 2006. – 80 с.
17. Рудаков К. В. Про алгебраїчної теорії універсальних і локальних обмежень для задач класифікації / К. В. Рудаков / Розпізнавання, класифікація, прогноз. Математичні методи та їх застосування. Вип. 1. – М.: Наука, 2007. – С. 176-200.
18. Тарасенко Ф. П. Введение в курс теории информации / Ф. П. Тарасенко. – Томск, изд. ТГУ, 1963. с. 236.
19. Федорук В. І. Теорія інформації та кодування / В. І. Федорук. – Чернівці, «Рута», 2001. с. 90.
20. Фу К. Структурні методи в розпізнаванні образів / К. Фу. – М.: Світ, 2005. – 144 с.
21. Цимбал В. П. Теория информации и кодирование / В. П. Цимбал. – Киев, «Вища школа», 1992. с. 263.
22. Цимбал В. П. Задачник по теории информации и кодирования / В. П. Цимбал. – Киев, «Вища школа», 1976. с. 275.
23. Яглом А. М. Вероятность и информация / А. М. Яглом, І. М. Яглом. – Москва, «Наука», 1973. с. 512.
24. Довбиш А. С. Основи теорії розпізнавання образів : навч. посіб. : у 2 ч. / А. С. Довбиш, І. В. Шелехов. – Суми : Сумський державний університет, 2015. – Ч. 1. – 109 с.
25. Алиев Т. А., Мусаева Н. Ф. Алгоритм уменьшения погрешностей оценки корреляционной функции сигнала с шумом. РАН СО «Автометрия», № 4, 1995, с.105-112.
26. Дмитриев В. И. Прикладная теория информации. – М.: Высшая школа., 1989 – 320 с.

НАВЧАЛЬНЕ ВИДАННЯ

Людмила Анатоліївна МАТВІЙЧУК

**ТЕОРЕТИЧНІ ОСНОВИ ІНФОРМАТИКИ
НАВЧАЛЬНИЙ ПОСІБНИК**

Комп'ютерна верстка – Л. А. Матвійчук

*Свідоцтво про державну реєстрацію
друкованого засобу масової інформації
серія КВ № 17500-6250 ПР від 16.11.2010 р.*

Підписано до друку ____ р. Формат 60 x 84 1/16.

Папір офсетний. Друк на різнографі.

Ум. друк. арк. ____ . Обл.-вид. арк. ____ .

Наклад 50 прим. Зам. № ____ .

Редакційно-видавничий відділ НУЧК імені Т.Г. Шевченка.

14013, вул. Гетьмана Полуботка, 53, к. 208.

Тел. 65-17-99.