Міністерство освіти і науки України
Сумський державний університет

4637 Методичні вказівки
до лабораторних робіт
із дисципліни **«Захист інформації в комп'ютерних системах»**
для студентів
спеціальності *«Електронні системи та компоненти»*
всіх форм навчання

Частина 1

Суми
Сумський державний університет
2019

Методичні вказівки до лабораторних робіт із дисципліни «Захист інформації в комп’ютерних системах» / укладачі: О. В. Бережна, Т. О. Протасова, О. В. Д’яченко. – Суми : Сумський державний університет, 2019 – Ч. 1. – 22 с.

Кафедра електроніки і комп’ютерної техніки

ЗМІСТ

ПЕРЕДМОВА

Інформаційна безпека − одна з головних проблем, з якою стикається сучасне суспільство. Причиною загострення цієї проблеми є широкомасштабне використання автоматизованих засобів накопичення, зберігання, оброблення та передавання інформації. Вирішення проблеми інформаційної безпеки пов'язане з гарантованим забезпеченням трьох її головних складових: доступності, цілісності і конфіденційності.

Під час виконання лабораторної роботи 1 передбачається розгляд статистичних принципів стиснення інформації з використанням методів Шеннона – Фано і Хаффмена.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Алгоритм Шеннона – Фано. Ефективне кодування методом Шеннона – Фано базується на основній теоремі Шеннона для каналу без перешкод.

Алгоритм Шеннона – Фано:

- літери алфавіту повідомлень виписуються в таблицю в порядку зменшення ймовірностей;
- літери алфавіту поділяються на дві групи так, щоб суми ймовірностей у кожній групі були наскільки можливо однаковими;
- всім літерам верхньої половини як перший символ приписується одиниця, а всім нижнім - нулі;
- кожна з отриманих груп зі свого боку розбивається на дві підгрупи з однаковими сумарними ймовірностями тощо, процес повторюється доти, поки в кожній підгрупі залишиться по одній букві.

Приклад 1. Розглянемо алфавіт із семи літер, приписавши кожній букві ймовірність її появи в повідомленні P (існують спеціальні таблиці ймовірностей появи букв різних алфавітів у повідомленнях, наприклад, табл. 1).

Таблиця 1 – Приклад кодування методом Шеннона – Фано

Літера	Ймовірність	Процес отримання кода	Код Шеннона
A	1/4	1 1 →	11
E	1/4	1 0 →	10
F	1/8	0 1 1 →	011
C	1/8	0 1 0 →	010
B	1/8	0 0 1 →	001
D	1/16	0 0 0 1 →	0001
G	1/16	0 0 0 0 →	0000

$$H(\xi) = \sum p_k \log_2 1/p_k.$$

$$H(\xi) = 2/4 + 2/4 + 3/8 + 3/8 + 3/8 + 4/16 + 4/16 = 2{,}625.$$

За умови звичайного двійкового кодування, не враховуючи статистичних характеристик, для представлення кожної букви потрібно три символи.

Найбільший ефект стиснення виходить у разі, коли ймовірності букв являють собою цілочисельні ступені двійки. Середня кількість символів на літеру в цьому разі точно дорівнює ентропії.

Середня кількість символів на літеру

$$L = \sum_{i=1}^{N} p_i n_i,$$

де p_i – ймовірність появи i-го символу алфавіту;

n_i, – кількість символів у кодовій комбінації i-го символу алфавіту.

Для розглянутого прикладу 1

$$L = \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 4 + \frac{1}{16} \cdot 4 = 2{,}625.$$

Різниця величин (L – H) – надмірність коду, а величина (L – H) / L – відносна надмірність.

Алгоритм Хаффмена. Суть алгоритму Хаффмена зводиться до такого:

- літери алфавіту повідомлень виписуються в основний стовпець таблиці в порядку зменшення ймовірностей;
- дві останні літери об'єднуються в одну допоміжну, якій приписується сумарна ймовірність;
- ймовірності букв, які не брали участі в об'єднанні, і отримана сумарна ймовірність знову розташовуються в порядку зменшення ймовірностей, а дві останні об'єднуються доти, поки не отримують єдину допоміжну літеру з ймовірністю, що дорівнює одиниці;
- далі для побудови коду використовується бінарне (кодове) дерево, у корені якого розташовується буква з ймовірністю одиниця, у разі розгалуженні гілки з більшою ймовірністю присвоюється код одиниця, а з меншою – код нуль (можливо, лівої – одиниця, а правої – нуль).

Приклад 2. Розглянемо умовний алфавіт із восьми літер, кожній із яких приписана відповідна ймовірність її появи в повідомленні (табл. 2).

Таблиця 2 – Приклад кодування методом Хаффмена

Літера	**Ймовір-ність**	**Допоміжні стовпці ймовірностей**							**Код Хаффмена**
Z1	0,22	0,22	0,22	0,26	0,32	0,42	**0,58**	**1**	01
Z2	0,20	0,20	0,20	0,22	0,26	**0,32**	**0,42**		00
Z3	0,16	0,16	0,16	0,20	**0,22**	**0,26**			110
Z4	0,16	0,16	0,16	**0,16**	**0,20**				111
Z5	0,10	0,10	**0,16**	**0,16**					100
Z6	0,10	**0,10**	**0,10**						1011
Z7	**0,04**	**0,06**							10101
Z8	**0,02**								10100

$L = 0{,}22 \cdot 2 + 0{,}20 \cdot 2 + 0{,}16 \cdot 3 + 0{,}16 \cdot 3 + 0{,}10 \cdot 3 + 0{,}10 \cdot 4 + + 0{,}04 \cdot 5 + 0{,}02 \cdot 5 = 2{,}8;$

$H = 2{,}76;$

$L - H = 0{,}04.$

Для порівняння: у коді Шеннона – Фано з таким самим розподілом ймовірностей $L - H = 0{,}08$.

Кодове дерево зображено на рис. 1.

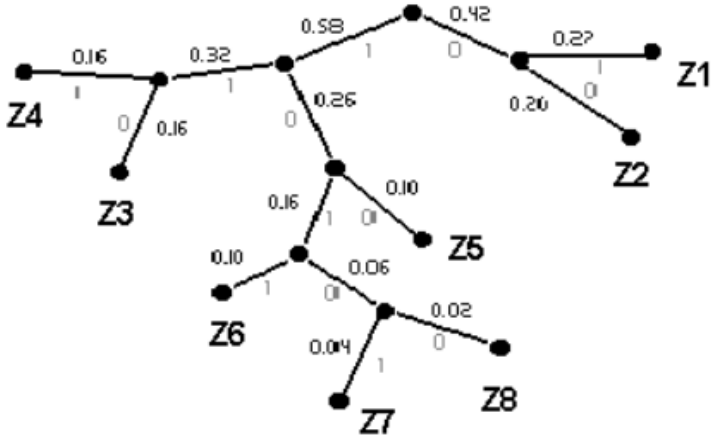


Рисунок 1 – Кодове дерево для прикладу 2

Для побудови кодів Шеннона – Фано і Хаффмена можна використовувати таблицю ймовірності окремих літер (табл. 3).

Завдання для самостійної роботи 1

1. Написати процедуру, що забезпечує ефективне кодування і підрахунок середньої кількості інформації на знак H; надмірності (L – H) і відносної надмірності одержуваного коду (L – H) / L. Для тестування програми використовувати два контрольних варіанта, наведених у табл. 4.

Таблиця 3 – Таблиця ймовірності окремих літер

Літера	Імовірність	Літера	Імовірність
_	0,175	Я	0,018
О	0,090	Ы	0,016
Е	0,072	З	0,016
А	0,062	Ъ, Ь	0,014
И	0,062	Б	0,014
Т	0,053	Г	0,013
Н	0,053	Ч	0,012
С	0,045	Й	0,010
Р	0,040	Х	0,009
В	0,038	Ж	0,007
Л	0,035	Ю	0,006
К	0,028	Ш	0,006
М	0,026	Ц	0,004
Д	0,025	Щ	0,003
П	0,023	Э	0,003
У	0,021	Ф	0,002

Таблиця 4 – Контрольні варіанти для завдання 1

Літера	Варіант 1		Варіант 2	
	Ймовірність	Код Шеннона	Ймовірність	Код Хаффмена
Z1	0,22	11	1/2	1
Z2	0,2	101	1/4	01
Z3	0,16	100	1/8	001
Z4	0,16	01	1/16	0001
Z5	0,1	001	1/32	00001
Z6	0,1	0001	1/64	000001
Z7	0,04	00001	1/128	0000001
Z8	0,02	00000	1/128	00000000
	H = 2,76; L = 2,84; L – H = 0,08		H = 163/64; L = 163/64; L – H = 0	

2. Джерело повідомлень породжує знаки A, B, C з вірогідністю 0,7; 0,2; 0,1.

Написати процедуру, що дозволяє:

- визначити ентропію джерела;
- вказати для цих трьох знаків оптимальне бінарне кодування і визначити середню довжину кодових комбінацій;
- закодувати всі пари AA, AB, ... (табл. 5);
- побудувати для цих дев'яти пар оптимальний бінарний код;
- збільшити блоковість коду до трисимвольних комбінацій і побудувати оптимальний бінарний код.

Таблиця 5 – Контрольні варіанти для завдання 2

Літера	**Ймовірність**	**Код**	**Знак**	**Ймовірність**	**Код пари**
A	0,7	0	AA	0,49	1
B	0,2	11	AB	0,14	011
C	0,1	10	BA	0,14	010
			AC	0,07	0011
			CA	0,07	0010
			BB	0,04	0001
			BC	0,02	00001
			CB	0,02	000001
			CC	0,01	000000
H = 1,157; L = 1,3; $L_{MAX} = 2,33L_{символов} = L_{MAX}/2 = 1,165$					

Зробити висновок про зміну надмірності коду зі збільшенням блоковості. Як це впливає на ефективність коду?

ЛАБОРАТОРНА РОБОТА 1

«Методи стиснення за Шенноном і Хафменом»

Мета роботи: ознайомлення зі статистичними принципами стиснення інформації з використанням методів Шеннона – Фано і Хаффмена.

Примітка. Для виконання лабораторної роботи на комп'ютері необхідно встановити файл Shannon-Huflman.exe, який міститься в архіві. Методи стиснення за Шенноном і Хаффменом.rar.

Описання лабораторної роботи. Робота виконується на персональному комп'ютері з використанням програми Shannon-Huffman.exe. Програма призначена для демонстрації методів стиснення інформації за алгоритмами Шеннона – Фано та Хаффмена (рис. 2).

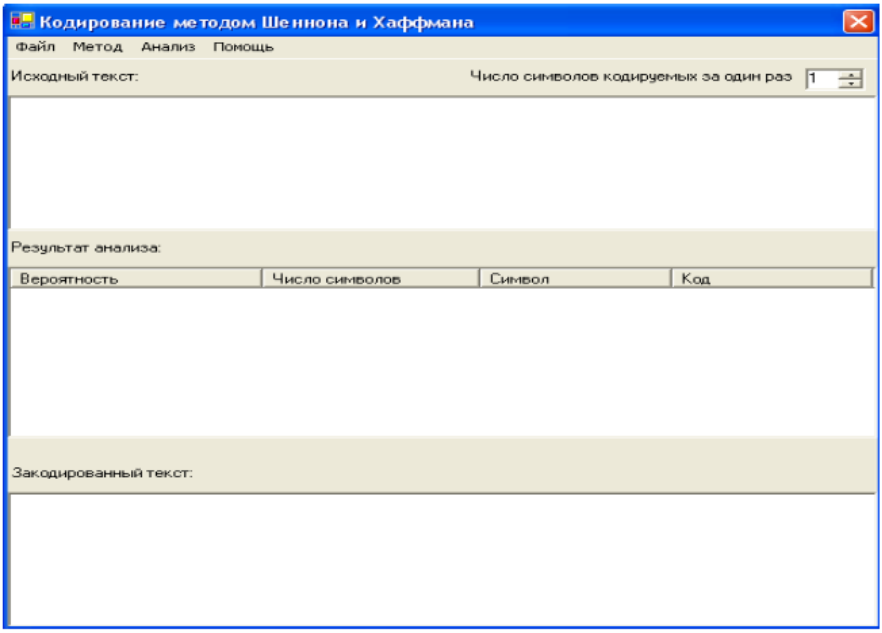


Рисунок 2 – Головне вікно програми

Для роботи з програмою користувач вибирає необхідний метод стиснення (закладка «Метод»).

У вікні програми ВИХІДНИЙ ТЕКСТ записується повідомлення довільної довжини (або завантажується повідомлення із заздалегідь підготовленого файлу). Потім необхідно вказати кількість символів, що кодуються за один раз.

Для підрахунку ймовірності появи букв у введеному повідомленні P, визначення ентропії джерела повідомлень H, середньої кількості символів у разі кодування однієї літери повідомлення I необхідно вибрати закладку АНАЛІЗ.

Для визначення ефективності кодування розраховується надмірність коду (L – H).

Коротко ознайомитися з відомостями про програму ви можете вибравши відповідну закладку (рис. 3).

Щоб вийти з програми, досить вибрати закладку «Файл» і далі «Вихід».

Приклад 3. Оригінал тексту аббввггггдддддеееее, що складається із символів алфавіту A = {а, б, в, г, д, е}, закодуємо методом Хаффмена.

Визначаємо ймовірність появи символу у вихідному повідомленні:

$$P = n/N,$$

де n – кількість повторів символу в повідомленні;

N – довжина повідомлення.

Впишемо у стовпець всі символи алфавіту в порядку зменшення ймовірностей їх появи в тексті (табл. 6).

Послідовно об’єднуючи два символи з найменшими ймовірностями появи символів у новий (об’єднаний) символ, ймовірність появи якого дорівнює сумі ймовірностей складових його символів, побудуємо дерево, кожен вузол якого має сумарну ймовірність всіх вузлів, що міститься нижче за нього. Дослідимо шлях до кожного листка дерева, позначаючи напрямок до кожного вузла (наприклад, направо – 0, наліво – 1) (рис. 4).

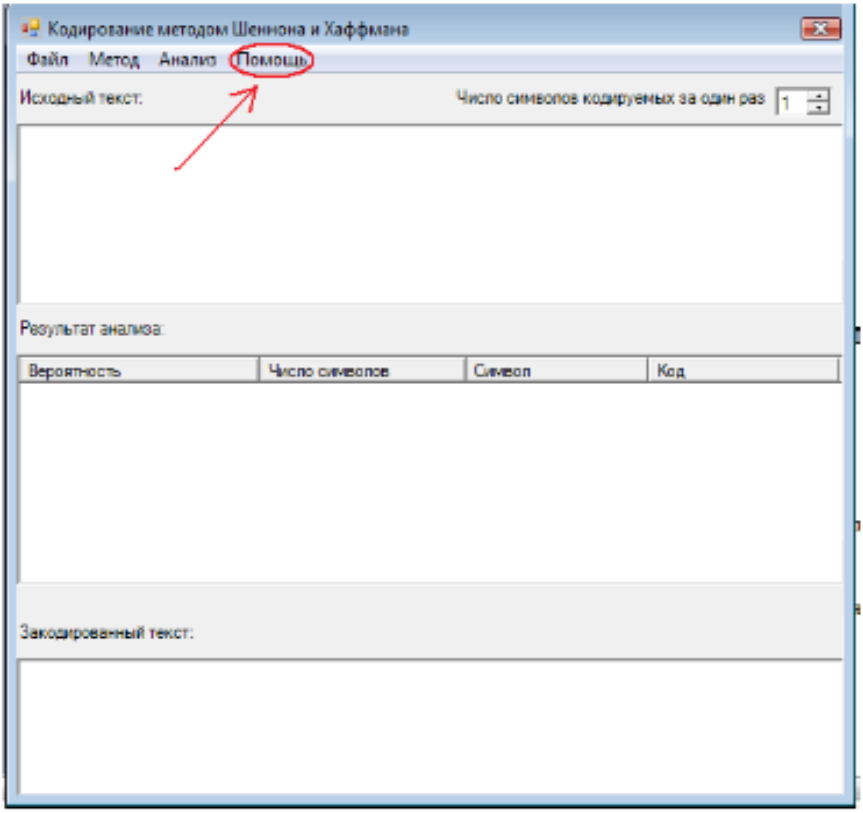


Рисунок 3 – Перегляд відомостей про програму

Таблиця 6 – Імовірність появи символів

Символ	**Кількість повторень символів у цьому повідомленні**	**Імовірність появи символів**
д	5	0,25
е	5	0,25
г	4	0,20
в	3	0,15
б	2	0,10
а	1	0,05

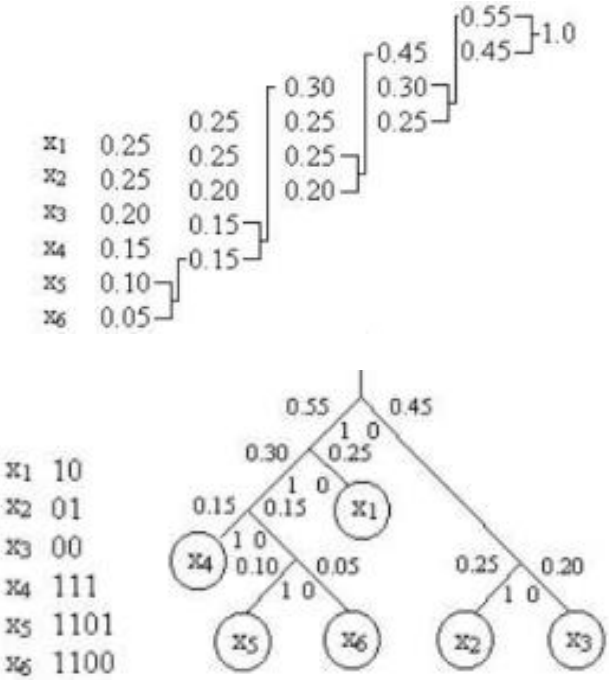


Рисунок 4 – Виконання прикладу 3

Середня кількість символів на букву повідомлення:

$$L = \sum_{i=1}^{n} P(i) \cdot n(i) = 2 \cdot 2 \cdot 0{,}25 + 2 \cdot 0{,}2 + 3 \cdot 0{,}15 + \\ + 4 \cdot (0{,}1 + 0{,}05) = 2{,}45,$$

де n (i) – кількість знаків у кодової комбінації *i*-го символу алфавіту;

P (i) – ймовірність появи *i*-го символу алфавіту.

Ентропія:

$$H = \sum_{i=1}^{n} P(i) \log \frac{1}{P(i)} = 2{,}425.$$

Значення надмірності коду: (L – H) = 0,025.

Порівнюємо отримані дані з результатами роботи програми (рис. 5).

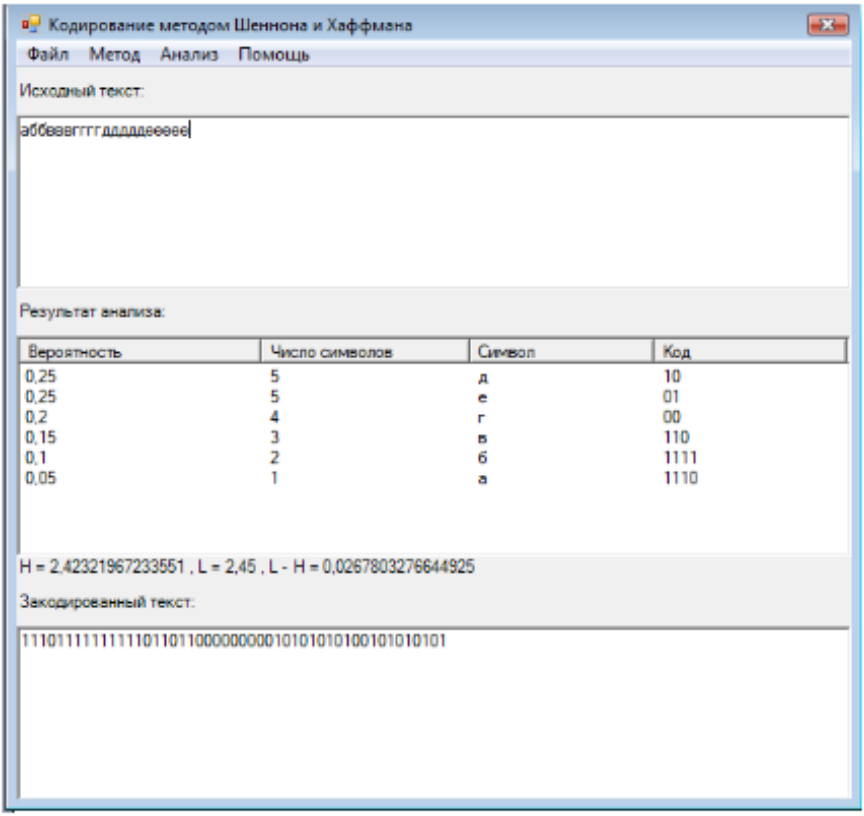


Рисунок 5 – Ілюстрація виконання процесу кодування методом Хаффмена

Приклад 4. Закодуємо вихідне методом Шеннона – Фано. Впишемо в стовпець всі символи в порядку зменшення ймовірності появи їх у тексті. Послідовно ділимо множину символів на дві підмножини так, щоб сума ймовірностей появи символів однієї підмножини приблизно дорівнювала сумі ймовірностей появи символів іншої. Для нижній підмножини кожному символу приписуємо 0, для верхній – 1. Подальші

розбиття повторюються доти, поки кожна підмножина не буде складатися з одного елемента (табл. 7).

Таблиця 7 –Кодування повідомлень методом Шеннона

Символ	**Кількість повторень символів у цьому повідомленні**	**Імовірність появи символів**	**Код Шеннона**		
Д	5	0,25	1	1	
Е	5	0,25	1	0	
Г	4	0,20	0	1	1
В	3	0,15	0	1	0
Б	2	0,10	0	0	1
А	1	0,05	0	0	0

Середня кількість символів на букву повідомлення:

$$L = \sum_{i=1}^{n} P(i)n(i) =$$

$$= 0{,}25 \cdot\ 2 + 0{,}25 \cdot\ 2 + 0{,}2 \cdot\ 3 + 0{,}15 \cdot\ 3 + 0{,}1 \cdot\ 3 + 0{,}25 \cdot\ 3 =$$
$$= 2{,}5.$$

Ентропія:

$$H = \sum_{i=1}^{n} P(i) \log \frac{1}{P(i)} = 2{,}425.$$

Значення надмірності коду: (L – H) = 0,075.

Порівнюємо отримані дані з результатами роботи програми (рис. 6).

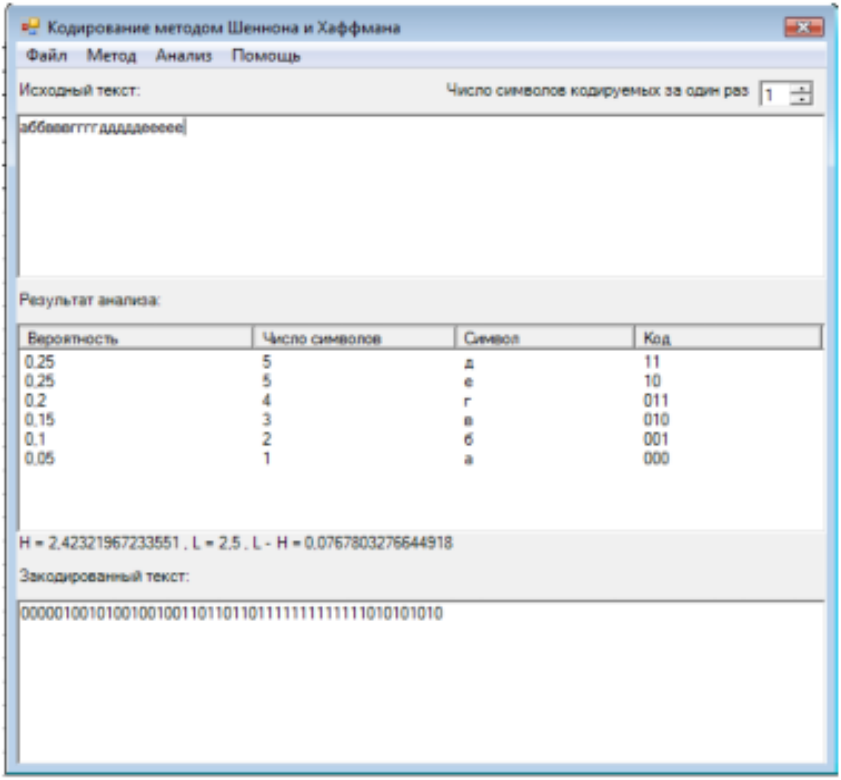


Рисунок 6 – Приклад кодування методом Шеннона

Завдання

1. Ознайомитися з відомостями про програму Shannon-Huffman.exe і розглянути приклади ефективного кодування інформації.

2. Запустити програму Shannon-Huffman.exe, призначену для демонстрації методів стиснення інформації з використанням алгоритмів Шеннона – Фано і Хаффмена.

3. Виконати такі завдання:

1) кількість символів алфавіту k = m (m – номер варіанта).

Скласти таке вихідне повідомлення, щоб:

а) символи алфавіту траплялися в повідомленні з однаковими ймовірностями,

б) символи алфавіту траплялися в повідомленні з різними ймовірностями;

2) ввести довільний зв'язний текст російською мовою. Це може бути прислів'я, вірш або довільний текст. Використовуючи результати роботи програми, необхідно проаналізувати алфавіт введеного повідомлення: підрахувати кількість символів алфавіту, значення ентропії H, середню кількість символів на знак L.

4. Зберегти у звіті екранні форми, що демонструють процес стиснення інформації.

5. Звіт із лабораторної роботи повинен містити результати аналізу програми Shannon-Huffman.exe під час кодування; алгоритм побудови коду Хаффмена і Шеннона із застосуванням знань із цієї тематики; результати порівняння різних методів кодування; висновки про ефективність застосованих методів стиснення.

6. Навести у звіті відповіді на контрольні питання відповідно до номера варіанта, зазначеного викладачем (табл. 8).

Додати до звіту про лабораторну роботу виконані завдання № 1-3.

Таблиця 8 – Контрольні питання

Номер варіанта	**Контрольні питання**
1, 5, 7, 3, 9, 18, 28	Які коди дозволяють виконувати однозначне декодування навіть без використання розділового символу? Наведіть приклади таких кодів
2, 4, 6, 8, 20, 22, 24, 26, 30	Назвіть умови побудови оптимальних кодів
11, 13, 15, 10, 17, 19, 27	З якою метою використовуються ефективні коди і які з них вам відомі?
12, 14, 16, 21, 23, 25, 29	Назвіть основні методи стиснення інформації без втрат

Варіанти завдань

Завдання 1. Повідомлення складається з послідовності двох букв А та В, ймовірності появи кожної з яких не залежать від того, яка була передана раніше, та дорівнюють 0,8 і 0,2 відповідно.

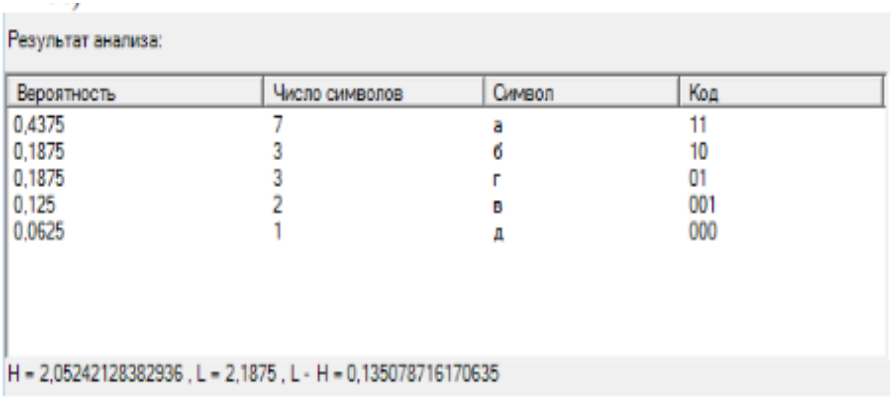
Провести кодування за методом Шеннона: а) окремих букв; б) блоків, що складаються з двобуквених сполучень; в) блоків, що складаються з трибуквених поєднань. Порівняти отримані коди за їхньою ефективністю.

Завдання 2. Скласти текст, який би відповідав даним, наведеним у табл. 9. Використовуючи програму Shannon-Huffman.exe, закодувати текст методом Хаффмена.

Таблиця 9 – Дані для виконання завдання 2

Номер варіанта	**Імовірність появи символів**	**Символ**	**Кількість символів**
1, 5, 7, 3, 9, 12, 14, 18, 28	0,333333333	О	2
	0,166666667	Г	1
	0,166666667	Р	1
	0,166666667	Д	1
	0,166666667	А	1
2, 4, 6, 8, 16, 21, 20, 22, 24, 30	0,25	Е	2
	0,25	Т	2
	0,125	О	1
	0,125	П	1
	0,125	Р	1
	0,125	А	1
11, 13, 15, 10, 17, 19, 23,25	0,25	Р	2
	0,25	А	2
	0,25	С	2
	0,125	Т	1
	0,125	Е	1

Завдання 3. Для варіантів а, б, в, наведених на рис. 6 і в табл. 10, скласти код Хаффмена. Розрахувати середню кількість символів на знак L; надмірність (L – H) і відносну надмірність отриманого коду (L – H)/L. Порівняти отримані значення з L, H, (L – H) для коду Шеннона, зробити висновки.

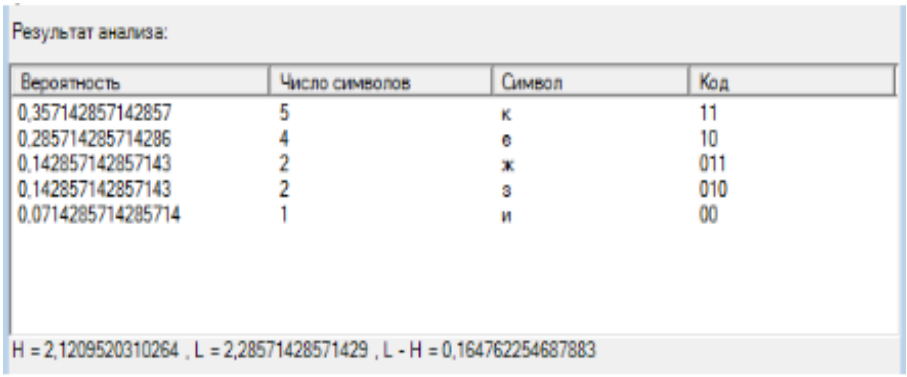


Результат анализа:

Вероятность	Число символов	Символ	Код
0,4375	7	а	11
0,1875	3	б	10
0,1875	3	г	01
0,125	2	в	001
0,0625	1	д	000

H = 2,05242128382936 , L = 2,1875 , L - H = 0,135078716170635

а)



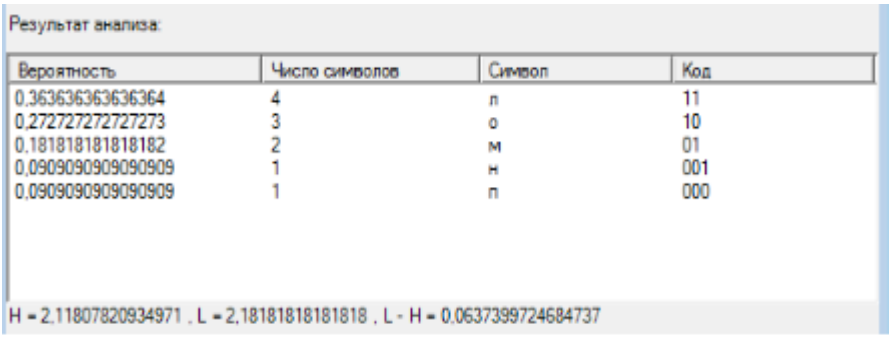
Результат анализа:

Вероятность	Число символов	Символ	Код
0,357142857142857	5	к	11
0,285714285714286	4	е	10
0,142857142857143	2	ж	011
0,142857142857143	2	з	010
0,0714285714285714	1	и	00

H = 2,1209520310264 , L = 2,28571428571429 , L - H = 0,164762254687883

б)

Рисунок 6 – Варіанти завдань


Результат анализа:

Вероятность	Число символов	Символ	Код
0,363636363636364	4	п	11
0,272727272727273	3	о	10
0,181818181818182	2	м	01
0,0909090909090909	1	н	001
0,0909090909090909	1	т	000

H = 2,11807820934971 , L = 2,18181818181818 , L - H = 0,0637399724684737

в)

Рисунок 6, аркуш 2

Таблиця 10 – Варіанти завдань

Номер варіанта	Завдання
1, 5, 7, 3, 9, 12, 14, 18, 28	3 а
2, 4, 6, 8, 20, 22, 23, 24, 25, 30	3 б
11, 13, 15, 16, 21, 10, 17, 19	3 в

СПИСОК ЛІТЕРАТУРИ

1. Информационная безопасность. Лабораторный практикум (+CD) : учебное пособие / А. В. Бабаш, Е. К. Баранова, Ю. Н. Мельников. – 2-е изд. – Москва : КНОРУС, 2013. – 136 с.

2. Баранова Е. К. Эффективное кодирование и защита информации : текст лекций для студентов специальности 510200. – Москва : МГУЛ, 2002. – 88 с.

3. Вернер М. Основы кодирования : учебник для вузов. – Москва : Техносфера, 2006. – 288 с.

4. Крушный В. В. Основы теории информации и кодирования. – Снежинск : СГФТА, 2005. – 69 с.

5. Морелос-Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение. – Москва : Техносфера, 2006. – 320 с.

6. Теория информации и кодирования. / Филоненков А. И., Самсонов Б. Б., Кречет Т. В., Плохов Е. М. – Москва : ФЕНИКС, 2002. – 288 с.

7. Хохлов Г. И. Основы теории информации : учеб. пособие для студ. высш. учеб. заведений / Г. И. Хохлов. – Москва : Академия, 2008. – 176 с.

8. Шеннон К. Математическая теория связи. Работы по теории информации и кибернетике / К. Шеннон. пер. с англ. ; под ред. Р. Л. Добрушина и О. Б. Лупанова. – Москва : ИЛ, 1963. – 512 с.

Навчальне видання

Методичні вказівки
до лабораторних робіт
із дисципліни **«Захист інформації в комп'ютерних системах»**
для студентів
спеціальності *«Електронні системи та компоненти»*
всіх форм навчання

Частина 1

Відповідальний за випуск А. С. Опанасюк
Редактор І. О. Кругляк
Комп'ютерне верстання О. В. Бережної

Формат 60x84/16. Ум. друк. арк. 1,40. Обл.-вид. арк. 1,32.

Видавець і виготовлювач
Сумський державний університет,
вул. Римського-Корсакова, 2, м. Суми, 40007
Свідоцтво суб'єкта видавничої справи ДК № 3062 від 17.12.2007.