Практична робота №4

Виконав: студент групи ПМІ-22 **Юрас Назар**

**Варіант-21**

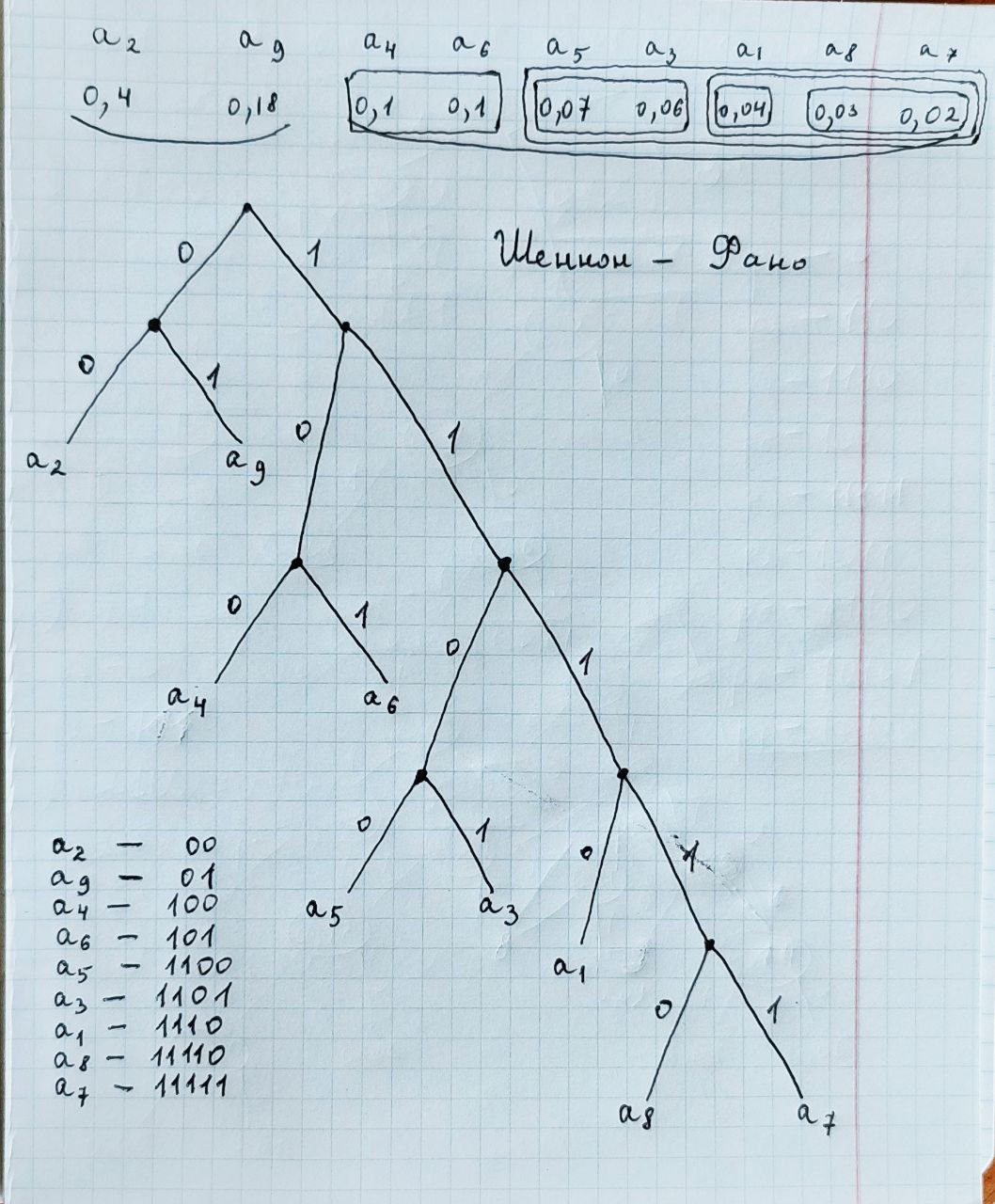
1. Значення ймовірностей pi, з якими дискретне джерело інформації генерує символи алфавіту, для різних варіантів наведені у табл. 4.1. Побудувати нерівномірні ефективні коди за алгоритмами Шеннона-Фано та Хаффмена. Порівняти ефективність кодів.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Варіант | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 |
| 21 | 0.04 | 0.4 | 0.06 | 0.1 | 0.07 | 0.1 | 0.02 | 0.03 | 0.18 |

**Код за алгоритмом Шеннона – Фано:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Символ** | **Імовірність** | **Код** | **Довжина** | **Імовірність \* Довжина** |
| **p2** | **0.4** | **00** | **2** | **0.8** |
| **p9** | **0.18** | **01** | **2** | **0.36** |
| **p4** | **0.1** | **100** | **3** | **0.3** |
| **p6** | **0.1** | **101** | **3** | **0.3** |
| **p5** | **0.07** | **1100** | **4** | **0.28** |
| **p3** | **0.06** | **1101** | **4** | **0.24** |
| **p1** | **0.04** | **1110** | **4** | **0.16** |
| **p8** | **0.03** | **11110** | **5** | **0.15** |
| **p7** | **0.02** | **11111** | **5** | **0.1** |

**Середня довжина: 2.69**

**Дерево Шеннона-Фано:**

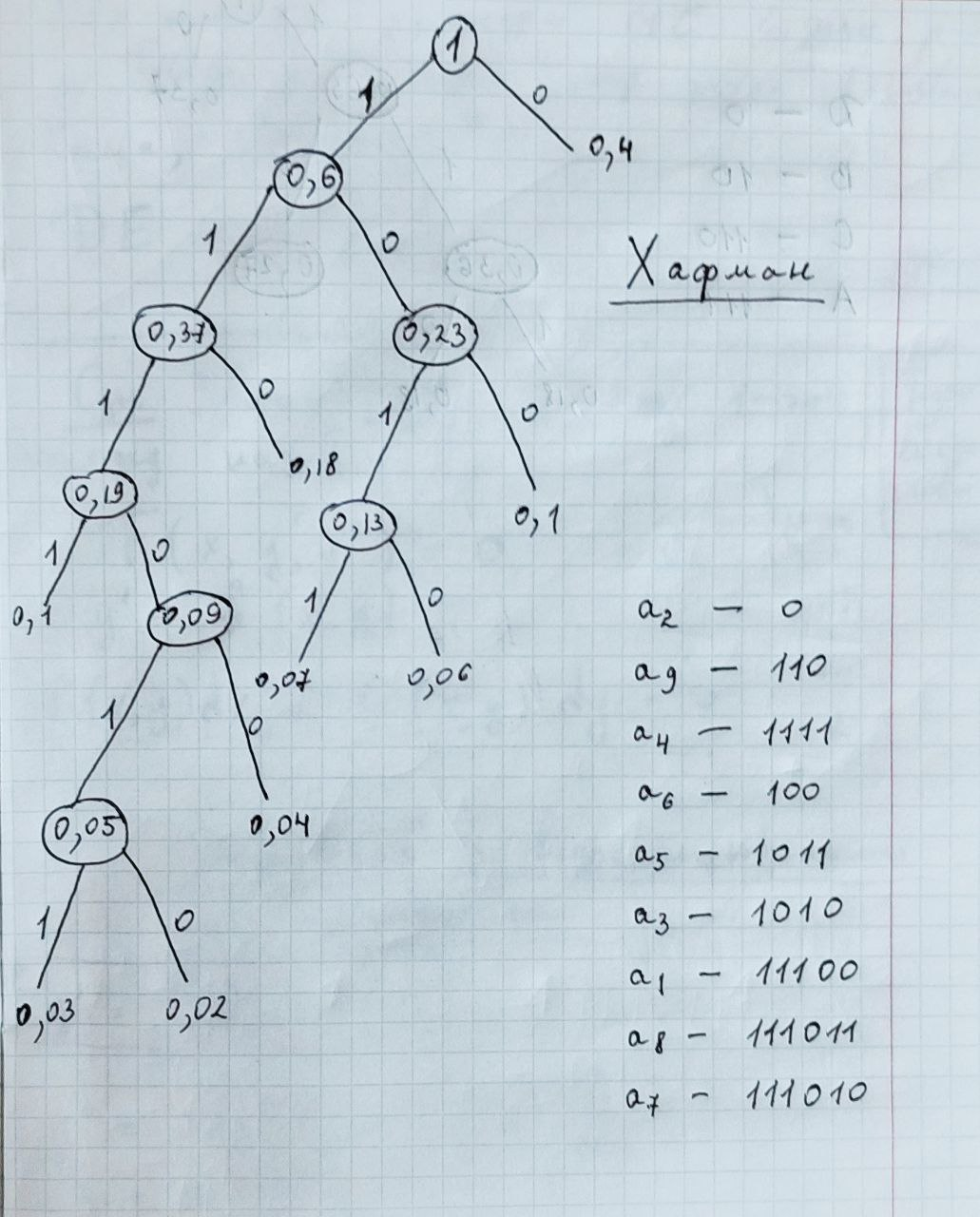
**H(X)** = - (0.4 log2 0.4 + 0.18 log2 0.18 + 0.1 log2 0.1 + 0.1 log2 0.1 + 0.07 log2 0.07 + 0.06 log2 0.06 + 0.04 log2 0.04 + 0.03 log2 0.03 + 0.02 log2 0.02) = **2.60095**

Відносна різниця середньої довжини оптимального ефективного коду та ентропії джерела становить ((2.69 – 2.60095) / 2.60095)\*100 = **3,4237%**

**Код за алгоритмом Хаффмана:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Символ** | **Імовірність** | **Код** | **Довжина** | **Імовірність \* Довжина** |
| **p2** | **0.4** | **0** | **1** | **0.4** |
| **p9** | **0.18** | **110** | **3** | **0.54** |
| **p4** | **0.1** | **1111** | **4** | **0.4** |
| **p6** | **0.1** | **100** | **3** | **0.3** |
| **p5** | **0.07** | **1011** | **4** | **0.28** |
| **p3** | **0.06** | **1010** | **4** | **0.24** |
| **p1** | **0.04** | **11100** | **5** | **0.2** |
| **p8** | **0.03** | **111011** | **6** | **0.18** |
| **p7** | **0.02** | **111010** | **6** | **0.12** |

**Середня довжина: 2.66**

**Дерево Хаффмана:**

Відносна різниця середньої довжини оптимального ефективного коду та ентропії джерела становить ((2.66 – 2.60095) / 2.60095)\*100 = **2.2703%**

1. Алфавіт дискретного джерела інформації складається з чотирьох символів X = {A,B,C,D}.Значення ймовірностей виникнення символів для різних варіантів наведені у табл. 4.2. Побудувати нерівномірні ефективні коди за алгоритмами Шеннона-Фано або Хаффмена для кодування поодиноких символів джерела та слів довжиною у два символи. Оцінити та порівняти ефективність отриманих кодів. Побудованими кодами закодувати фрагмент повідомлення довжиною у 30 символів, що був згенерований джерелом.

**p(A) = 0.18, p(B) = 0.27, p(C) = 0.18, p(D) = 0.37,**

AABCDCDDBBABABBCACAAADABCBCABC

Алгоритм Хафмана:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi \* li |
| D | 0.37 | 0 | 1 | 0.37 |
| B | 0.27 | 10 | 2 | 0.54 |
| C | 0.18 | 110 | 3 | 0.54 |
| A | 0.18 | 111 | 3 | 0.54 |

**Середня довжина**: 1.99

H(X) = - (0.37 log2 0.37 + 0.27 log2 0.27 + 0.18 log2 0.18 + 0.18 log2 0.18) = **1.93136**

Відносна різниця середньої довжини оптимального ефективного коду та ентропії джерела становить ((1.99 – 1.93136) / 1.93136)\*100 = **3,0362%**

A A B C D C D D B B A B A B B C A C A A A D A B C B C A B C

111 111 10 110 0 110 0 0 10 10 111 10 111 10 10 110 111 110 111 111 111 0 111 10 110 10 110 111 10 110

**73 символи**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Символ** | **Імовірність** | **Код** | **Довжина** | **pi \* li** |
| **DD** | **0.1369** | **110** | **3** | **0,4107** |
| **BD** | **0,0999** | **010** | **3** | **0,2997** |
| **DB** | **0,0999** | **001** | **3** | **0,2997** |
| **BB** | **0,0729** | **1110** | **4** | **0,2916** |
| **CD** | **0,0666** | **1010** | **4** | **0,2664** |
| **DC** | **0,0666** | **1011** | **4** | **0,2664** |
| **AD** | **0,0666** | **1000** | **4** | **0,2664** |
| **DA** | **0,0666** | **1001** | **4** | **0,2664** |
| **BC** | **0,0486** | **0000** | **4** | **0,1944** |
| **CB** | **0,0486** | **0001** | **4** | **0,1944** |
| **BA** | **0,0486** | **11110** | **5** | **0,243** |
| **AB** | **0,0486** | **11111** | **5** | **0,243** |
| **CC** | **0,0324** | **01110** | **5** | **0,162** |
| **AC** | **0,0324** | **01111** | **5** | **0,162** |
| **CA** | **0,0324** | **01100** | **5** | **0,162** |
| **AA** | **0,0324** | **01101** | **5** | **0,162** |

**Середня довжина: 3,8901**

Закодований фрагмент довжиною 2:

AA BC DC DD BB AB AB BC AC AA AD AB CB CA BC

**01101 0000 1011 110 1110 11111 11111 0000 01111 01101 1000 11111 0001 01100 0000**

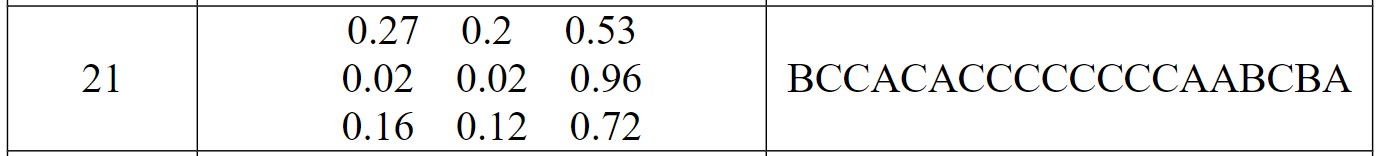
Довжина: 66 символів

H(X) = **3.86273**

Відносна різниця середньої довжини оптимального ефективного коду та

ентропії джерела дорівнює ((3,8901 - 3,86273)/(3,86273))\*100 = **0.7085%**

1. Алфавіт марковського дискретного джерела інформації, що має глибину пам’яті h =1, складається з трьох символів: X = {A,B,C}. Значення умовних ймовірностей виникнення символів для різних варіантів наведені у другому стовпчику табл. 4.3.1. Побудувати нерівномірні ефективні коди за алгоритмом Шеннона-Фано або Хаффмена для кодування поодиноких символів джерела та слів довжиною у два символи.2. Побудувати марковський алгоритм для кодування символів джерела.3. Оцінити та порівняти ефективність отриманих кодів та марковського алгоритму.4. Побудованими кодами закодувати фрагмент повідомлення (наведений у третьому стовпчику табл. 4.3) довжиною у 20 символів, що був зґенерований джерелом.

****

**Символ А:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi \* li |
| С | 0.53 | 1 | 1 | 0.53 |
| А | 0.27 | 01 | 2 | 0.54 |
| В | 0.2 | 00 | 2 | 0.4 |

**Середня довжина: 1.47**

**Символ В:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi \* li |
| С | 0.96 | 1 | 1 | 0.96 |
| В | 0.02 | 01 | 2 | 0.04 |
| А | 0.02 | 00 | 2 | 0.04 |

**Середня довжина: 1.04**

**Символ С:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi \* li |
| С | 0.72 | 1 | 1 | 0.72 |
| А | 0.16 | 01 | 2 | 0.32 |
| В | 0.12 | 00 | 2 | 0.24 |

**Середня довжина: 1.28**

P(A)\*p(A|A) + p(B)\*p(A|B) + p(C)\*p(A|C) = 0

P(B) \* p(B|A) + p(B)\*p(B|B) + p(C)\*p(B|C) = 0

P(A) + p(B) + p(C) = 0

-0.73 p(A) + 0.02 p(B) + 0.16 p(C) = 0

0.2 p(A) – 0.98 p(B) + 0.12 p(C) = 0

p(A) + p(B) + p(C) = 1

lcep = lcep/A \* p(A) + lcep/B \* p(B) + lcep/C \* p(C)

p(A) = **0,1608**, p(B) = **0,1208**, p(C) = **0,7184**

lcep = 0.1608 \* 1.47 + 0.1208 \* 1.04 + 0.7184 \* 1.28 = 0,236376 + 0,125632 + 0,919552 = **1,28156**

H(X|A) = - ( 0.27 log2 0.27 + 0.2 log2 0.2 + 0.53 log2 0.53) = 1.45985

H(X|B) = - ( 0.02 log2 0.02 + 0.02 log2 0.02 + 0.96 log2 0.96) = 0.28229

H(X|C) = - ( 0.16 log2 0.16 + 0.12 log2 0.12 + 0.72 log2 0.72) = 1.13131

H(X) = 0.1608 \* 1.45985 + 0.1208 \* 0.28229 + 0.7184 \* 1.13131 = 0,23474388 + 0,034100632 + 0,812733104 = **1,08158**

Відносна різниця середньої довжини оптимального ефективного коду та

ентропії джерела дорівнює ((1.28156–1.08158)/(1.08158))\*100 = **18,4896%**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi\*li |
| C | 0.7184 | 0 | 1 | 0.7184 |
| A | 0.1608 | 01 | 2 | 0.3216 |
| B | 0.1208 | 00 | 2 | 0.2416 |

**Середня довжина: 1,2816**

H(X) = 1.1351

Закодований фрагмент:

B C C A C A C C C C C C C C A A B C B A

00 0 0 01 0 01 0 0 0 0 0 0 0 0 01 01 00 0 00 01

28 символів

Відносна різниця середньої довжини оптимального ефективного коду та

ентропії джерела дорівнює ((1.2816–1.1351)/(1.1351))\*100 = **12,9063%**

За Хаффманом:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Символ | Імовірність | Код | Довжина | pi\*li |
| 1. СС | 0,517248 | 1 | 1 | 0.517248 |
| 3) CA | 0,114944 | 001 | 3 | 0,344832 |
| 5) AC | 0,085224 | 0111 | 4 | 0,340896 |
| 1. BC | 0,115968 | 010 | 3 | 0,347904 |
| 4) CB | 0,086208 | 000 | 3 | 0,258624 |
| 6) AA | 0,043416 | 01101 | 5 | 0,21708 |
| 7) AB | 0,03216 | 011001 | 6 | 0,19296 |
| 9) BA | 0,002416 | 0110000 | 7 | 0,016912 |
| 8) BB | 0,002416 | 0110001 | 7 | 0,016912 |

**Середня довжина: 2.2534**

H(X) = 2.2167

Відносна різниця середньої довжини оптимального ефективного коду та

ентропії джерела дорівнює ((2.2534–2.2167)/(2.2167))\*100 = **1,6556%**

**Закодований фрагмент довжиною 2 символи:**

BC CA CA CC CC CC CC AA BC BA

010 001  001 1 1 1 1 01101 010 0110000

28 символів

Найефективнішим виявився метод Хаффмана, тоді марківський, а тоді Фано