Міністерство освіти і науки України

Черкаський державний технологічний університет

Кафедра програмного забезпечення автоматизованих систем

**ЗВІТ**

з лабораторної роботи №2

з предмету «Основи інтеграції інформаційних потоків»

|  |  |
| --- | --- |
| Перевірив:  Д. т. н., проф., зав. каф. ПЗАС  Первунінський С.М.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2020 р. | Виконав:  студент 1-го курсу  групи МПЗ-1904  Гаврилюк В. Є. |

Черкаси 2020

**Лабораторна робота №2**

**Тема роботи:** Побудова ефективного коду по методу Хаффмена.

**Мета роботи:** Закріпити теоретичні знання і набути навички по ефективному кодуванню інформації по методиці Хаффмена.

**Завдання:**

1. Користуючись методикою Хаффмена, побудувати ефективний ОНК для кодування букв алфавіту української мови для значення основи коду а = 5. Ймовірність p(i) букв узяти з файлу D:\text1.txt.
2. Визначити середню тривалість кодових комбінацій одержаного коду. Обчислити коефіцієнти стиснення.
3. Користуючись штучним прийомом Мічеля В.С., побудувати ефективний бінарний код для кодування букв алфавіту української мови.

**Теоретичні відомості**

Розглянемо методику побудови ОНК, запропоновану Хаффменом.

Задача побудови ОНК підстави ***а*** для некорельованих алфавітів m формулюється таким чином: серед всіх можливих кодів підстави *а* без розділових знаків, що володіють властивістю префікса (ніяка коротша кодова комбінація не є початком (префіксом) довшої кодової комбінації), знайти код, для якого мінімально можлива середня тривалість кодової комбінації

, (1)

де  - апріорна вірогідність передачі (появи) - го символу алфавіту m;  - тривалість - ї кодової комбінації в коді з підставою *а*. Як показав Хаффмен, для того, щоб даний префіксний код забезпечував мінімально можливе значення величини , а отже був би ефективним, необхідно і достатньо виконання наступних трьох умов:

1. Якщо виписати символи у порядку убиваючої вірогідності:

, де , то тривалості відповідних кодових комбінацій повинні задовольняти співвідношенню .

1. В усякому разі, дві останні, але не більше ніж **,** кодові комбінації, рівні по тривалості і відрізняються значенням тільки останнього кодового знаку: де  , де . Кожна можлива послідовність кодових знаків повинна або сама бути кодовою комбінацією, або мати своїм префіксом використовувану кодову комбінацію.

Відповідно до цих трьох умов методика побудови ОНК полягає в наступному.

Символи алфавіту ***m*** записуються у порядку убиваючої вірогідності в основний стовпець. Останні по вірогідності  символів, де - таке найбільше число з інтервалу , що  - ціле число, об'єднуються в новий символ, вірогідність якого дорівнює сумарній вірогідності символів, його складових. ** символів, що залишились, та одержаний новий символ знову виписуються в перший додатковий стовпець у порядку убиваючої вірогідності. Останні **а** по вірогідності символи першого додаткового стовпця знову об'єднуються в допоміжний символ.  символів, що залишилися та знову одержаний допоміжний символ також виписуються у порядку убиваючої вірогідності і т.д. до тих пір, поки **а** останніх символів додаткового стовпця не дадуть допоміжний символ з вірогідністю, рівній одиниці.

Ключ побудови коду полягає в наступному. Число кодових знаків кодової комбінації *i* - го символу, записуваної справа наліво (від кінця), рівне числу об'єднань, в яких бере участь даний символ на шляху утворення допоміжного символу з одиничною вірогідністю.

Якщо номерам символів в об'єднаннях довільно зіставити **а** знаків [0, 1, 2,..., (а-1)], то значення *N* - го кодового знаку, відлічуваного від кінця кодової комбінації, визначається номером, займаним даним символом (або допоміжним символом, в утворенні якого він брав участь) в його *N* - у об'єднанні від початку побудови.

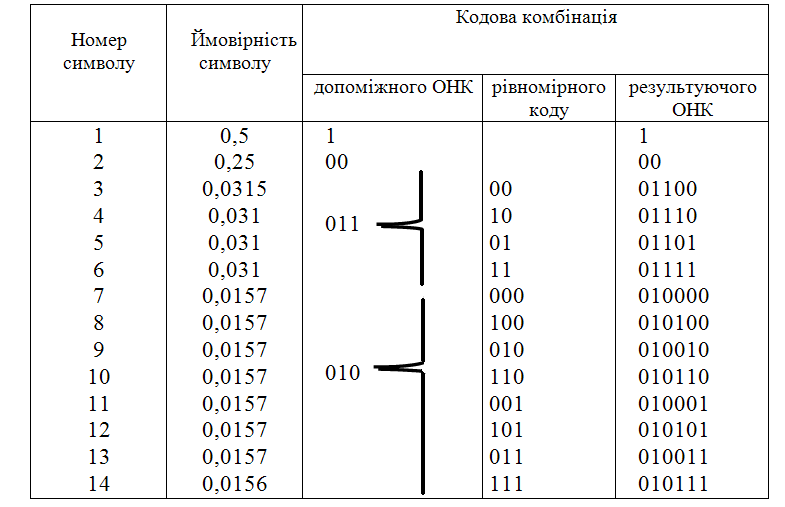
Приклад побудови деяких ОНК по даній методиці приведений в табл.1. Слід зазначити, що сформульовані Хаффменом три умови ОНК в явному вигляді не містять властивості префікса. Проте властивість префікса автоматично виходить із запропонованої Хаффменом методики побудови кодових комбінацій ОНК і прийнятого напряму запису кодових комбінацій справа наліво.

Необхідно відзначити, що при великому початковому алфавіті **m** безпосереднє застосування методики Хаффмена для складання ОНК приводить до громіздких побудов. У цих випадках іноді зручно скористатися відомим штучним прийомом, запропонованим Мічелем В.С. (Michel W.S.) [2]. Виписаний у порядку убиваючої вірогідності алфавіт **m** розбивається на значно менше число груп символів **m1**<**m.** При цьому бажано, щоб число об'єднуваних в групи символів складало цілий ступінь основи коду  **(****,****,…)**а вірогідність складових групи символів була б по можливості рівна між собою. Вірогідність такої групи є сумарна вірогідність складових її символів початкового алфавіту**m1.** За допомогою звичної методики Хаффмена будується допоміжний ОНК для меншого алфавіту. Кодова комбінація даного допоміжного ОНК є префіксом кодової комбінацій шуканого ОНК для символів, що становлять дану групу. Як доповнення (суфіксів) до цього префікса приписується кодова комбінація рівномірного коду на n розрядів, де **аn** - число символів в даній групі. Приклад такої побудови ОНК приведений в табл.2.

Таблиця 1



Таблиця 2



**Хід роботи**

Таблиця 3 – Методика Хаффмена

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Символ** | **Вірогідність** | **Code** | **A(i)** | **A(i) × p(i)** |
| 1 | о | 0,1036 | 001 | 3 | 0,3108 |
| 2 | н | 0,0826 | 1110 | 4 | 0,3304 |
| 3 | а | 0,0721 | 1100 | 4 | 0,2884 |
| 4 | и | 0,0609 | 1001 | 4 | 0,2436 |
| 5 | в | 0,0567 | 0111 | 4 | 0,2268 |
| 6 | т | 0,0553 | 0110 | 4 | 0,2212 |
| 7 | і | 0,0552 | 0101 | 4 | 0,2208 |
| 8 | е | 0,0503 | 0001 | 4 | 0,2012 |
| 9 | р | 0,0491 | 0000 | 4 | 0,1964 |
| 10 | д | 0,0393 | 11011 | 5 | 0,1965 |
| 11 | к | 0,0378 | 11010 | 5 | 0,189 |
| 12 | с | 0,0347 | 10111 | 5 | 0,1735 |
| 13 | л | 0,0331 | 10110 | 5 | 0,1655 |
| 14 | м | 0,0322 | 10101 | 5 | 0,161 |
| 15 | я | 0,0285 | 10000 | 5 | 0,1425 |
| 16 | у | 0,0284 | 01001 | 5 | 0,142 |
| 17 | п | 0,0252 | 111111 | 6 | 0,1512 |
| 18 | з | 0,0219 | 111101 | 6 | 0,1314 |
| 19 | б | 0,015 | 101000 | 6 | 0,09 |
| 20 | г | 0,0149 | 100011 | 6 | 0,0894 |
| 21 | ь | 0,0147 | 100010 | 6 | 0,0882 |
| 22 | х | 0,0119 | 1111101 | 7 | 0,0833 |
| 23 | й | 0,0115 | 1111001 | 7 | 0,0805 |
| 24 | ч | 0,0115 | 1111100 | 7 | 0,0805 |
| 25 | ц | 0,0099 | 1111000 | 7 | 0,0693 |
| 26 | ю | 0,0076 | 1010010 | 7 | 0,0532 |
| 27 | ж | 0,0069 | 0100011 | 7 | 0,0483 |
| 28 | є | 0,0069 | 0100010 | 7 | 0,0483 |
| 29 | ш | 0,0066 | 0100001 | 7 | 0,0462 |
| 30 | ф | 0,0064 | 0100000 | 7 | 0,0448 |
| 31 | ї | 0,0054 | 10100111 | 8 | 0,0432 |
| 32 | щ | 0,0029 | 101001101 | 9 | 0,0261 |
| 33 | ' | 0,0008 | 101001100 | 9 | 0,0072 |

Таблиця 4 - Прийомом, запропонованим Мічелем

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № | Символ | Вірогідність | Сумарна вірогідність | Допоміжний ОНК | Рівномірного коду | Результуючий ОНК |
| 1 | о | 0,1036 | 0,1036 | 1 | 0 | 10 |
| 2 | н | 0,0826 | 0,1547 | 110 | 0 | 1100 |
| 3 | а | 0,0721 | 1 | 1101 |
| 4 | и | 0,0609 | 0,1176 | 100 | 0 | 1000 |
| 5 | в | 0,0567 | 1 | 1001 |
| 6 | т | 0,0553 | 0,1105 | 10 | 0 | 100 |
| 7 | і | 0,0552 | 1 | 101 |
| 8 | е | 0,0503 | 0,0994 | 0 | 0 | 00 |
| 9 | р | 0,0491 | 1 | 01 |
| 10 | д | 0,0393 | 0,1449 | 101 | 00 | 10100 |
| 11 | к | 0,0378 | 01 | 10101 |
| 12 | с | 0,0347 | 10 | 10110 |
| 13 | л | 0,0331 | 11 | 10111 |
| 14 | м | 0,0322 | 0,1143 | 11 | 00 | 11001 |
| 15 | я | 0,0285 | 01 | 1101 |
| 16 | у | 0,0284 | 10 | 1110 |
| 17 | п | 0,0252 | 11 | 1111 |
| 18 | з | 0,0219 | 0,155 | 111 | 0000 | 1110000 |
| 19 | б | 0,015 | 0001 | 1110001 |
| 20 | г | 0,0149 | 0010 | 1110010 |
| 21 | ь | 0,0147 | 0011 | 1110011 |
| 22 | х | 0,0119 | 0100 | 1110100 |
| 23 | й | 0,0115 | 0101 | 1110101 |
| 24 | ч | 0,0115 | 0110 | 1110110 |
| 25 | ц | 0,0099 | 0111 | 1110111 |
| 26 | ю | 0,0076 | 1000 | 1111000 |
| 27 | ж | 0,0069 | 1001 | 1111001 |
| 28 | є | 0,0069 | 1010 | 1111010 |
| 29 | ш | 0,0066 | 1011 | 1111011 |
| 30 | ф | 0,0064 | 1100 | 1111100 |
| 31 | ї | 0,0054 | 1101 | 1111101 |
| 32 | щ | 0,0029 | 1110 | 1111110 |
| 33 | ' | 0,001 | 1111 | 1111111 |

**Висновок**: на даній лабораторній роботі я побудував ефективний код по методу Хафмана. Закріпив теоретичні знання і набув навички по ефективному кодуванню інформації по методиці Хаффмена. Побудував таблиці ефективного кодуванню інформації.