**Міністерство освіти і науки України**  
**Національний технічний університет України**  
**“Київський політехнічний інститут”**

Інститут Прикладного системного аналізу  
Кафедра Системного проектування

Лабораторна робота з дисципліни  
«Теорія інформації та кодування»

на тему  
**«СТИСНЕННЯ ДАНИХ В ДИСКРЕТНИХ ПОВІДОМЛЕННЯХ З ВИКОРИСТАННЯМ АЛГОРИТМУ ГАФФМАНА»**

Виконав:

студент 2 курсу

групи ДА-92

Насікан Д. Ю.

Київ 2021

**Мета роботи -** вивчення алгоритмів ефективного кодування без втрат та їх реалізації для стиснення даних в дискретних повідомленнях з використанням алгоритму Гаффмана, оцінка надмірності повідомлень, розробка програми для дослідження ефективності кодування даних у файлах різних типів і довільної довжини.

1. **ЗАВДАННЯ**
2. Ознайомитися з методикою ефективного кодування Гаффмана.
3. Використовуючи табличний процесор MS Excel закодуйте алгоритмом Гаффмана текстове повідомлення з вашим прізвищем ім'ям, по батькові, датою і місцем народження. Підрахуйте ентропію текстового повідомлення, середню довжину кодових комбінацій закодованого повідомлення, коефіцієнт стиснення і оцініть ефективність кодування. Перевірте префіксність коду і розкодуйте повідомлення.
4. Розробити програму на мові програмування високого рівня для кодування і декодування текстових файлів довільної довжини. При вводі текстових даних передбачити два режиму вводу кодів: введення з клавіатури, а також введення даних з текстового файлу з використанням однобайтових кодових таблиць. За допомогою розробленої програми перевірити дані, отримані при виконанні п.2.
5. За допомогою розробленої при виконанні п.3 програми дослідити залежність коефіцієнту стиснення від довжини текстового файлу. Результати представити в табличній формі.
6. Модіфікувати програму, розроблену при виконанні п.4, для стиснення файлів довільних типів даних за алгоритмом Гаффмана. Провести дослідження ефективності розробленої програми шляхом порівняння результатів стиснення текстових, мультимедійних і exe –файлів з допомогою розробленої програми і архіватора WinZip . Результати занести в таблицю.
7. **ХІД РОБОТИ**
   1. Використовуючи табличний процесор MS Excel закодуємо методом Шеннона-Фано текстове повідомлення з вашим прізвищем ім'ям, по-батькові, датою і місцем народження:

|  |  |  |
| --- | --- | --- |
| **Символ** | **Імовірність появи** | **Код** |
| **a** | **0.116** | **011** |
| **\_** | **0.093** | **1111** |
| **н** | **0.07** | **1010** |
| **2** | **0.07** | **1100** |
| **0** | **0.07** | **1011** |
| **с** | **0.047** | **10011** |
| **і** | **0.047** | **0010** |
| **и** | **0.047** | **1000** |
| **т** | **0.047** | **0001** |
| **р** | **0.047** | **11101** |
| **о** | **0.047** | **11100** |
| **в** | **0.047** | **0100** |
| **.** | **0.047** | **0101** |
| **1** | **0.047** | **0011** |
| **к** | **0.023** | **110100** |
| **д** | **0.023** | **110101** |
| **м** | **0.023** | **00000** |
| **ю** | **0.023** | **00001** |
| **й** | **0.023** | **10010** |
| **ч** | **0.023** | **110110** |
| **з** | **0.023** | **110111** |

Повідомлення: насікан дмитро юрійович 22.10.2001 заставна

Код: 1010011100110010110100011101011111101010000010000001111011110011110000111101001010010111000100100011011011111100110001010011101101011100101110110011111111011101110011000101101001010011

Ентропія текстового повідомлення: 4.22

Середня довжина кодових комбінацій - 4.6.

З використанням однобайтових таблиць 1 символ повідомлення кодувався 8 бітами. Після кодування методом Хаффмана 1 символ повідомлення кодується 4.6 бітами. Отже, коеф. стиснення – 1.74. Як бачимо, кодування досить ефективне, так як дозволяє використовувати коди, довжина яких майже в 2 рази менша.

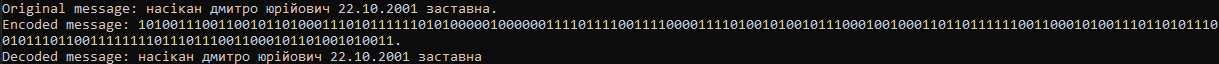
Проаналізувавши таблицю кодів бачимо, що жодна кодова послідовність, яка кодує якийсь символ не є префіксом більш довгої комбінації. Отже, даний код є префіксним.

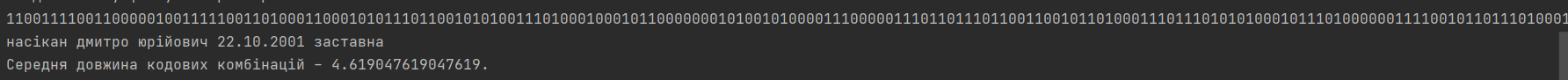
Після розкодування повідомлення отримуємо повідомлення: насікан дмитро юрійович 22.10.2001 заставна.

* 1. Розробимо програму на мові програмування високого рівня для кодування і декодування текстових файлів довільної довжини:

Лістинг програми додається окремим файлом, так як код має значний розмір.

Перевіримо результати пункту 2.1 на правильність, користуючись програмою:





Отриманий код:

1010011100110010110100011101011111101010000010000001111011110011110000111101001010010111000100100011011011111100110001010011101101011100101110110011111111011101110011000101101001010011

Розкодоване повідомлення: насікан дмитро юрійович 22.10.2001 заставна

Середня довжина кодових комбінацій: 4.619047619047619.

* 1. За допомогою розробленої при виконанні п.2.2 програми дослідимо залежність коефіцієнту стиснення від довжини текстового файлу:

|  |  |  |
| --- | --- | --- |
| **Довжина текстового файлу** | **Коефіцієнт стиснення** | **Середня довжина кодових комбінацій** |
| **43** | **1.73** | **4.61** |
| **196** | **1.68** | **6.075** |
| **619** | **1.64** | **6.85** |
| **1024** | **1.63** | **7.41** |
| **1694** | **1.58** | **7.69** |
| **4493** | **1.58** | **8.8** |

Як бачимо, що більша довжина файлу, то меншим стає коефіцієнт стиснення.

* 1. Модіфікуємо програму, розроблену при виконанні п.4, для стиснення файлів довільних типів даних за алгоритмом Шеннона-Фано:

Протестуємо програму та зведемо дані в таблицю порівняння:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **вхідний файл** | | | **Результат роботи розробленої програми** | | | **Архіватор WinRAR** | | |
| **Назва** | **Розмір файлу** | **Ентропія** | **Розмір файлу** | **Ентропія** | **Коефіцієнт стиснення** | **Розмір файлу** | **Ентропія** | **Коефіцієнт стиснення** |
| **1.png** | **3 KB** | **7.8743** | **3 KB** | **7.9246** | **1** | **1 KB** | **7.871** | **1** |
| **2.jpg** | **147 KB** | **7.9847** | **148 KB** | **7.99** | **0.99** | **147 KB** | **7.99** | **1** |
| **3.exe** | **1376 KB** | **7.887** | **1361 KB** | **7.983** | **1.011** | **1376 KB** | **7.88** | **1** |
| **4.pdf** | **9284 KB** | **7.672** | **8951 KB** | **7.94** | **1.0372** | **6697 KB** | **7.99** | **1.38** |

1. **ВИСНОВКИ**

У ході даної лабораторної роботи було розглянуто метод кодування Хаффмана, та розроблена програма, що реалізує цей алгоритм кодування для текстових файлів. Як приклад, було взято повідомлення «насікан дмитро юрійович 22.10.2001 заставна», у результаті перекодування, отримали коефіцієнт стиснення 1.73.

Для тестування програми було взято файли різної довжини, як можна помітити з таблиці, коефцієнт стиснення майже не залежить від довжини файлу.

У другій частині роботи програма була модифікованою під усі типи файлів. Тепер файл зчитувався у бінарному режимі, і методом Хаффмана кодувалися байти повідомлення. Як бачимо з порівняльної таблиці, алгоритм найкраще стискає файли формати exe.