**Imię i nazwisko:** Jakub Pranica **Grupa:** Informatyka -13i **Data:** 24-04-2017 **Przedmiot:** Metody Programowania

**Wykładowca:** dr inż. Zbigniew Kokosiński **Semestr:** 2

**Temat:** Napisać algorytm kodujący podany przez użytkownika tekst za pomocą algorytmu Huffmana.

**Kodowanie Huffmana** to jedna z najprostszych i łatwych w implementacji metod kompresji bezstratnej. Została opracowana w 1952 roku przez Amerykanina Davida Huffmana. Algorytm ten wykorzystywany jest w wielu profesjonalnych metodach kompresji tekstu, obrazów i dźwięków, również w połączeniu z innymi metodami. Kodowanie Huffmana stanowi więc kanon kompresji i jest adaptowane dla każdego tekstu.

**Opis programu:** Program wczytuje tekst z wcześniej przygotowanego pliku lub prosi użytkownika o wpisanie go z klawiatury. Wczytany tekst zostaje przeanalizowany przez program pod kątem ilości wystąpień każdego ze znaków (tworzony jest zbiór występujących symboli oraz zbiór stowarzyszonych z nim prawdopodobieństw). Na podstawie uzyskanych danych program rozpoczyna generację tak zwanego "drzewa Huffmana", z którego następnie odczytuje ciągi bitowe. Powstają w ten sposób słowa kodowe, których długość jest odwrotnie proporcjonalna do prawdopodobieństwa odpowiadającego im znaku. Kompresja polega na zastąpieniu symboli otrzymanymi kodami - program wypisuje na ekran lub do pliku wczytany wcześniej tekst w formie zakodowanej.

**Napotkane problemy: 1.**Zapis drzewa i jego węzłów za pomocą struktur w języku programowania c++ **2.**Wczytywanie dużego tekstu z pliku i analiza zawartych w nim znaków. **3.**Obliczenie oraz przypisanie odpowiednich prawdopodobieństw do każdego z występujących znaków **4.**Realizacja algorytmu Huffmana i wypisanie na ekran lub do pliku wyników programu.

**Przykładowy wynik programu:** Program wczytuje tekst z pliku lub z klawiatury i koduje go za pomocą algorytmuHuffmana.

**(Przykład 1)** Skąd pobrać dane? 1-wczytaj z pliku 2-pozwol wpisać z klawiatury podać **1**

**Wczytałem następujący tekst z pliku text.txt:**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam scelerisque just

o vitae condimentum gravida. Aenean efficitur leo aliquet auctor tempus.

**Znak: Ilość wystąpień**

**: 19**

**,: 1**

**.: 3**

**A: 1**

**L: 1**

**N: 1**

**a: 9**

**c: 7**

**d: 4**

**e: 16**

**f: 2**

**g: 2**

**i: 13**

**j: 1**

**l: 7**

**m: 7**

**n: 6**

**o: 8**

**p: 3**

**q: 2**

**r: 7**

**s: 8**

**t: 12**

**u: 10**

**v: 2**

**Uzyskane kody dla poszczególnych znaków:**

**m: 0000**

**r: 0001**

**e: 001**

**o: 0100**

**s: 0101**

**d: 01100**

**g: 011010**

**q: 011011**

**a: 0111**

**: 100**

**v: 101000**

**,: 1010010**

**A: 1010011**

**L: 1010100**

**N: 1010101**

**.: 101011**

**u: 1011**

**t: 1100**

**n: 11010**

**p: 110110**

**j: 1101110**

**f: 1101111**

**i: 1110**

**c: 11110**

**l: 11111**

**Wczytany tekst w formie zakodowanej:**

10101000100000100100001001110110110010110110000100011000100111110100000110001011

11011001000111000000111001010010100111100100110100101001111101100001110010110001

10001110110011101101101110010111110111011010011010100001111111110110010101110010

10101101111111111110111000010001011111000111111001000111100101011011101100110011

01110101101011100010010010100011101100011100110011110010011010011001110000000111

01011001011000010001101000010111101000111001100011110101110010100110011101000101

11110101000011101111110111111101111011101100101100011001111100101001000111111111

11001101110110011100100011110111111011000100000110011000010000110110101101011010

11

**(Przykład 2)** Skąd pobrać dane? 1-wczytaj z pliku 2-pozwol wpisać z klawiatury podać **2**

**Wpisz tekst do zakodowania:** Ala ma kota.

**Znak: Ilość wystąpień**

: 2

.: 1

A: 1

a: 3

k: 1

l: 1

m: 1

o: 1

t: 1

**Uzyskane kody dla poszczególnych znaków:**

m: 000

o: 001

a: 01

t: 100

: 101

.: 1100

A: 1101

k: 1110

l: 1111

**Wczytany tekst w formie zakodowanej:**

1101111101101000011011110001100011100

Wyjście programu zapisano również do pliku: wynik.txt

**Pseudokod:**

**Wypisz na ekran:** "Program wczytuje tekst z pliku lub z klawiatury i koduje go za pomocą algorytmu Huffmana. \n"

**Wypisz na ekran:** "Skąd pobrać dane? 1-wczytaj z pliku 2-pozwol wpisać z klawiatury podać \n"

**Jeżeli użytkownik wpisał 1, wczytaj tekst z pliku - w przeciwnym wypadku poproś użytkownika o ręczne wpisanie, a następnie zapisz go w zmiennej string text.**

**Utwórz tablicę: ccount[]**

**Policz ilość wystąpień każdego znaku:**

for(int i = text.size() - 1; i >= 0; i--) ccount[text[i]]++;

**Utwórz strukturę wezel reprezentującą węzeł drzewa binarnego:**

struct wezel{

wezel \* rodzic, \* lewo, \* prawo;

char znak;

int wartosc;

};

**Utwórz listę tab i wypełnij ją początkowymi węzłami, po czym rozpocznij tworzenie drzewa binarnego Huffmana:**

**Dopóki rozmiar tab większy od 1** (tab.size() > 1) **wykonuj w pętli:**

**Wyszukaj na liście dwa najmniejsze elementy (w1 oraz w2)**

**Z węzłów zawartych w w1 i w2 utwórz nowy węzeł w:** w->rodzic = NULL; w->lewo = w1; w->prawo = w2;

w->lewo->rodzic = w->prawo->rodzic = w; w->wartosc = w1->wartosc + w2->wartosc;

**Dodaj w na listę, usuwając z niej jednocześnie w1 oraz w2**

**Wszystkie utworzone po drodze węzły tworzą drzewo binarne (Huffmana). Wywołaj funkcje przeszukaj(), która przejdzie przez całe drzewo wypisując kody Huffmana oraz przypisując je do poszczególnych znaków za pomocą tablicy kodyH.**

**Odczytaj ponownie każdy znak wczytanego tekstu wypisując przy tym na ekran lub do pliku odpowiadający mu kod Huffmana:**

for(int i = 0; i < text.size(); i++){ cout << kodyH[text[i]]; }

**Wnioski:** Na podstawie analizy wyników zwracanych przez program i obliczeń wykonanych poza nim, można stwierdzić, iż program działa i prawidłowo posługuje się algorytmem Huffmana, w celu zakodowania podanego przez użytkownika tekstu. Kodowanie Huffmana opiera sie o częstość występowania znaków, gdzie długości ciągów kodowych dobierane są do statystyki źródła wiadomości. Redukcja wielkości danych przy stosowaniu tego algorytmu wynosi ok. 50 %.

**Literatura:**

* *Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein*: **Wprowadzenie do algorytmów.** Warszawa: Wydawnictwa Naukowo-Techniczne, 2003.
* *Zbigniew Kokosiński* **wykład i prezentacja nr. 7** Metody programowania, Politechnika Krakowska 2017.
* **pl.wikipedia.org/wiki/Kodowanie\_Huffmana**