

Teoria Informacji - sprawozdanie

Lab 5: Kompresja bezstratna - wstęp

Marek Subocz 136806

L3, czwartek 15:10

February 6, 2021

1 Zadania teoretyczne

Jaka jest najkrótsza możliwa długość kodu o stałej długości dla korpusu podanego na zajęciach?

Jest to długość sześciu bitów (37 znaków, 5 i $\log_2(37)$ i 6).

Ile wyniesie stopień kompresji?

$8 \text{ bitów} / 6 \text{ bitów} = 1,333$

2 Zadanie praktycznie

2.1 Środowisko

Do wykonania zadania użyto języka Python3. Jediną biblioteką spoza standardowych jest bitarray.

2.2 Wyniki

```
florida from complications of alzheimer s disease he was 71
{' ': '000000', 'e': '000001', 'a': '000010', 't': '000011', 'i': '000100', 'n': '000101',
'o': '000110', 'r': '000111', 's': '001000', 'h': '001001', 'l': '001010', 'd': '001011',
'c': '001100', 'm': '001101', 'u': '001110', 'f': '001111', 'p': '010000', 'g': '010001',
'b': '010010', 'w': '010011', 'y': '010100', 'v': '010101', 'k': '010110', '1': '010111',
'0': '011000', '9': '011001', '2': '011010', 'j': '011011', '8': '011100', '3': '011101',
'5': '011110', 'x': '011111', '4': '100000', '7': '100001', '6': '100010', 'z': '100011',
'q': '100100'}
florida from complications of alzheimer s disease he was 71
Tekst oryginalny i odkodowany są takie same.
```

Jaki jest rozmiar oryginalnego pliku, a jaki pliku (lub plików) z kodem i zakodowaną reprezentacją?

- 10,8 MB - oryginalny plik
- 8,1 MB - plik z zakodowaną reprezentacją
- $555 \text{ B} = 0.000555 \text{ MB}$ - plik z kodem

Stopień kompresji wyniósł: $10,8 \text{ MB} / 8,1 \text{ MB} = 1,333$

3 Wnioski

Wielkość pliku zmniejszyła się, ponieważ zaimplementowane kodowanie umożliwiało użycie mniejszej ilości znaków niż oryginalne, wobec tego możliwa była bezstratna kompresja.

Co można zrobić żeby bardziej skompresować tekst?

Można zastosować kompresję ze zmienną długością kodu.

Co z nieużyтыми kodami?

Oznaczają one, że tekst można bardziej skompresować również za pomocą zmiany symboli użytych do kodowania, ponieważ kod nie jest idealnie efektywny.

Jak odkodowywać kody o zmiennej długości (ang. variable-length code)?

Trzeba upewnić się, że żadne słowo kodowe nie jest prefiksem innego słowa kodowego.