

# Laboratorium 3

## Kodowanie Huffmana

Jan Rajczyk

23 kwietnia 2021

### 1 Wstęp

Celem tego laboratorium było zaimplementowanie algorytmu kompresji Huffmana, określenie tego, jak dobrze dokonywana jest kompresja oraz zmierzenie czasu kompresji i dekompresji dla różnego rodzaju tekstów.

Pliki, na których były wykonywane testy podzieliłem na trzy kategorie:

#### 1. Pliki tekstowe

- **Tekst 1:** *Krzyżacy* w języku angielskim o rozmiarze 1.2 MB
- **Tekst 2:** Plik tekstowy o rozmiarze 195 kB
- **Tekst 3:** Plik tekstowy o rozmiarze 11 kB
- **Tekst 4:** Plik tekstowy o rozmiarze 3 kB

#### 2. Pliki źródłowe

- **Plik 1:** Plik źródłowy o rozmiarze 1.1 MB
- **Plik 2:** Plik źródłowy o rozmiarze 89 kB
- **Plik 3:** Plik źródłowy o rozmiarze 10 kB
- **Plik 4:** Plik źródłowy o rozmiarze 2 kB

#### 3. Pliki ze znakami losowanymi z rozkładu jednostajnego

- **Plik 1:** Plik o rozmiarze 1 MB
- **Plik 2:** Plik o rozmiarze 98 kB
- **Plik 3:** Plik o rozmiarze 20 kB
- **Plik 4:** Plik o rozmiarze 1 kB

## 2 Statyczne kodowanie Huffmana

Struktury, z których korzystałem to głównie *bitarray* oraz *Dict*. Wyniki przeprowadzonych testów zostały przedstawione w tabelach:

### 1. Pliki tekstowe

tekst	współczynnik kompresji	czas kompresji [s]	czas dekompresji [s]
Tekst 1	0.44343111529572343	0.09365990000000002	2.1303619
Tekst 2	0.4373396944134046	0.01312610000000003	0.18940609999999998
Tekst 3	0.4305668198612238	0.0017905999999996425	0.011293799999999798
Tekst 4	0.3564614050303556	0.0005657000000001133	0.000656400000000014

Tablica 1: Wyniki dla statycznego kodowania Huffmana dla plików tekstowych

### 2. Pliki źródłowe

tekst	współczynnik kompresji	czas kompresji [s]	czas dekompresji [s]
Plik 1	0.32261616306528085	0.09513170000000004	1.6510794999999998
Plik 2	0.3681079768684059	0.005256299999999214	0.13161760000000022
Plik 3	0.32852677220933746	0.0015410000000004587	0.011380499999999572
Plik 4	0.25420439844760667	0.0006716000000004385	0.001588300000000764

Tablica 2: Wyniki dla statycznego kodowania Huffmana dla plików źródłowych

### 3. Pliki ze znakami losowanymi z rozkładu jednostajnego

tekst	współczynnik kompresji	czas kompresji [s]	czas dekompresji [s]
Plik 1	-0.08800000000000008	0.0008014999999996775	0.002840299999999907
Plik 2	0.15785000000000005	0.002146500000000051	0.06709009999999993
Plik 3	0.16706	0.007285699999999505	0.33922890000000017
Plik 4	0.16856099999999996	0.06509650000000011	3.4366039000000006

Tablica 3: Wyniki dla statycznego kodowania Huffmana dla plików ze znakami losowanymi z rozkładu jednostajnego

**Wnioski:** Widzimy, że współczynnik kompresji waha się między 0.25 a 0.45 przy plikach źródłowych i tekstowych, przy czym przeważnie jest on wyższy dla plików tekstowych. Jeżeli chodzi o pliki tworzone ze znaków losowanych z rozkładu jednostajnego to widać, że dla małych plików współczynnik kompresji jest nawet ujemny, dla większych już natomiast jest on na stałym poziomie około 0.17.

Jeżeli chodzi o czasy działania to algorytm kompresji jest dużo szybszy od algorytmu dekompresji.