

Mikołaj Bujok
Szymon Miękus
Grupa BSI-2

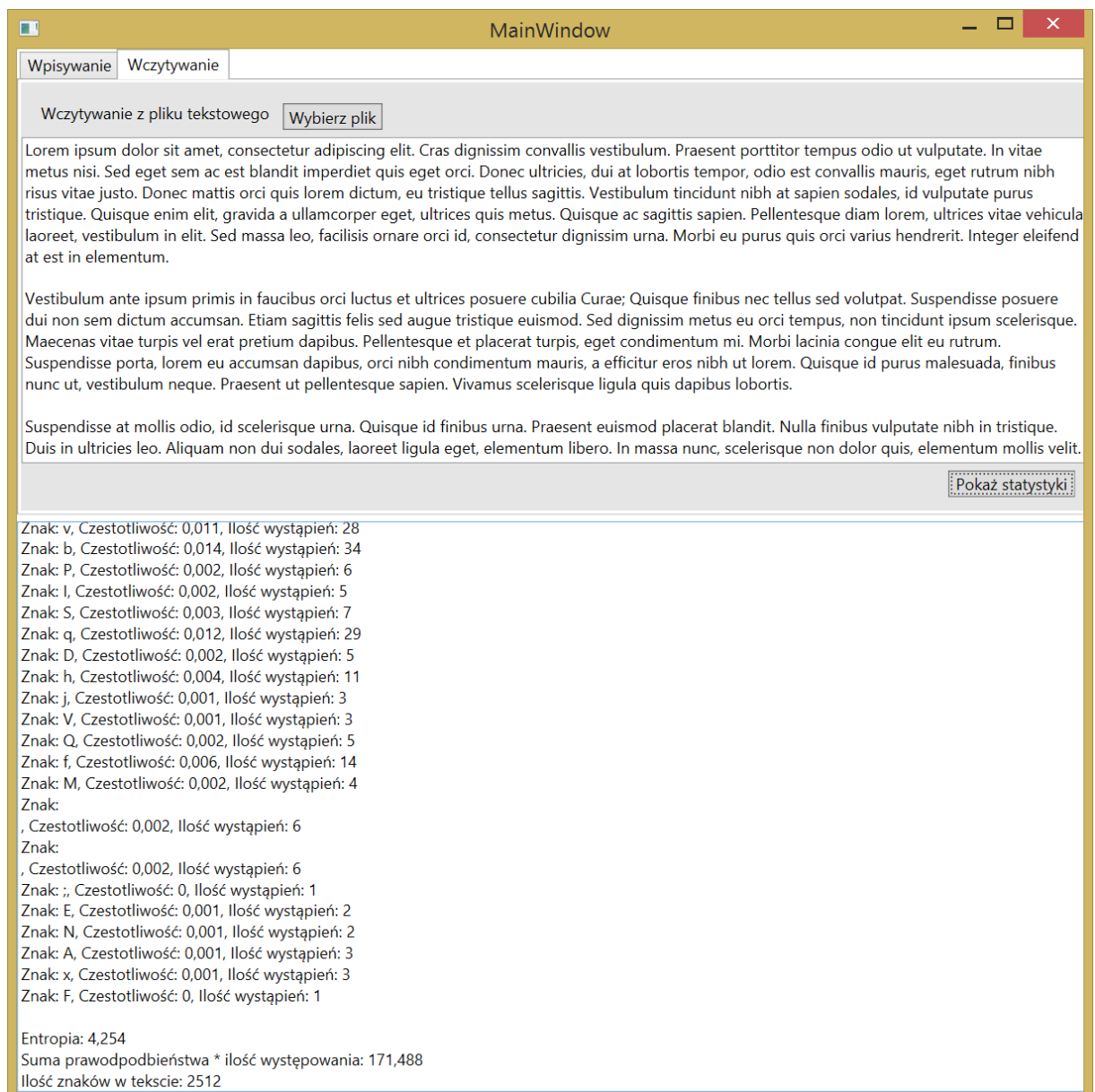
Sprawozdanie z zajęć „Teoria informacji i kodowanie”

1. Temat zajęć

Na poprzednich zajęciach laboratoryjnych badaliśmy stopień kompresji poszczególnych programów komercyjnych na trzech różnej wielkości plikach tekstowych. Porównywaliśmy go do sytuacji idealnej.

2. Przebieg i rezultaty

Zaimplementowaliśmy dodatkowe funkcje w naszej aplikacji okienkowej z poprzednich zajęć laboratoryjnych. Zawierały one funkcje pozwalające na wypisanie liczby wszystkich liter w tekście, oraz prawdopodobieństwo ich wystąpienia.



Rysunek 1: Zrzut ekranu aplikacji

Po stworzeniu trzech plików tekstowych (małego, średniego i dużego) przeprowadziliśmy badanie ich właściwości. Następnie skompresowaliśmy te pliki za pomocą wbudowanego do systemu Windows narzędzia do kompresji zip, oraz za pomocą programu PeaZip do formatu 7z.

Po zebraniu danych oraz wykonaniu niezbędnych obliczeń uzyskaliśmy wyniki przedstawione w tabeli poniżej.

Plik	Rozmiar w bitach dla			ilość znaków	entropia	Rozmiar/ilość znaków	
	txt	zip	7z			zip	7z
mały	176	1088	1112	22	3,88	49,451	50,54
średni	20096	9320	9728	2512	4,254	3,17	3,87
duży	577056	161704	141656	72132	4,262	2,242	1,96

3. Wnioski

Dla plików o bardzo małych rozmiarach kompresja jest niezalecana, ponieważ narzut dodatkowych danych związanych z kompresją przewyższa wielkość pliku przed kompresją. W przypadku plików o dużych rozmiarach sytuacja okazała się być nawet lepsza od idealnej, to mogło być spowodowane faktem użycia słów wygenerowanych przez „Lorem ipsum”, które powtarzały się cyklicznie w kompresowanym pliku tekstowym. W normalnym pliku przy jego większych rozmiarach narzutu danych związanych z kompresją jest znikomy, a więc kompresja jest opłacalna. Nie ma występuje również sytuacja lepsza od idealnego rozwiązania.