



# Dokumentacja projektu Kodowanie Shannona-Fano (pary znaków)

## **Teoria informacji i kodowanie** Informatyka, sem. VII

Damian Filipowicz  
Krzysztof Łuczak  
Maciej Marciniak  
Dawid Wiktorski  
BSI-2

## I. Opis zadania

Zadanie obejmuje utworzenie programu kompresującego pliki z użyciem algorytmu Shannona-Fano dla pary znaków.

W algorytmie wyróżnia się 6 etapów:

1. Określenie prawdopodobieństwa występowania wszystkich par znaków.
2. Sortowanie listy par znaków według prawdopodobieństw.
3. Podział grup par znaków na dwie części o możliwie równej sumie wag par znaków.
4. Przyporządkowanie jednej grupie par znaków binarne "0", a drugiej binarne "1".
5. Powtórzenie dla każdej podgrupy o zbiorze par znaków większej niż 1 kroków 3. i 4.
6. Przypisanie par znaków odpowiednich kodów wynikających z wcześniejszych podziałów.

## II. Wykorzystane technologie

Moduł odpowiedzialny za wygenerowanie słownika kodowego

- Język: *Python 3.5*
- System operacyjny: *Windows*
- Środowisko programistyczne (IDE): *Sublime Text 3*

Aplikacja desktopowa

- Język: *C# 7.0*
- Rodzaj aplikacji: *Windows Forms Application*
- Biblioteka graficzna: *Windows Forms*
- System operacyjny: *Windows*
- Środowisko programistyczne (IDE): *Visual Studio 2017*

### III. Podział prac

Rozdział przedstawia podział prac nad zadaniem projektowym między członkami zespołu znajduje się w Tab. 1. Podział obejmuje podzadania oraz zadania wspólne.

*Tab.1. Podział prac nad zadaniem projektowym*

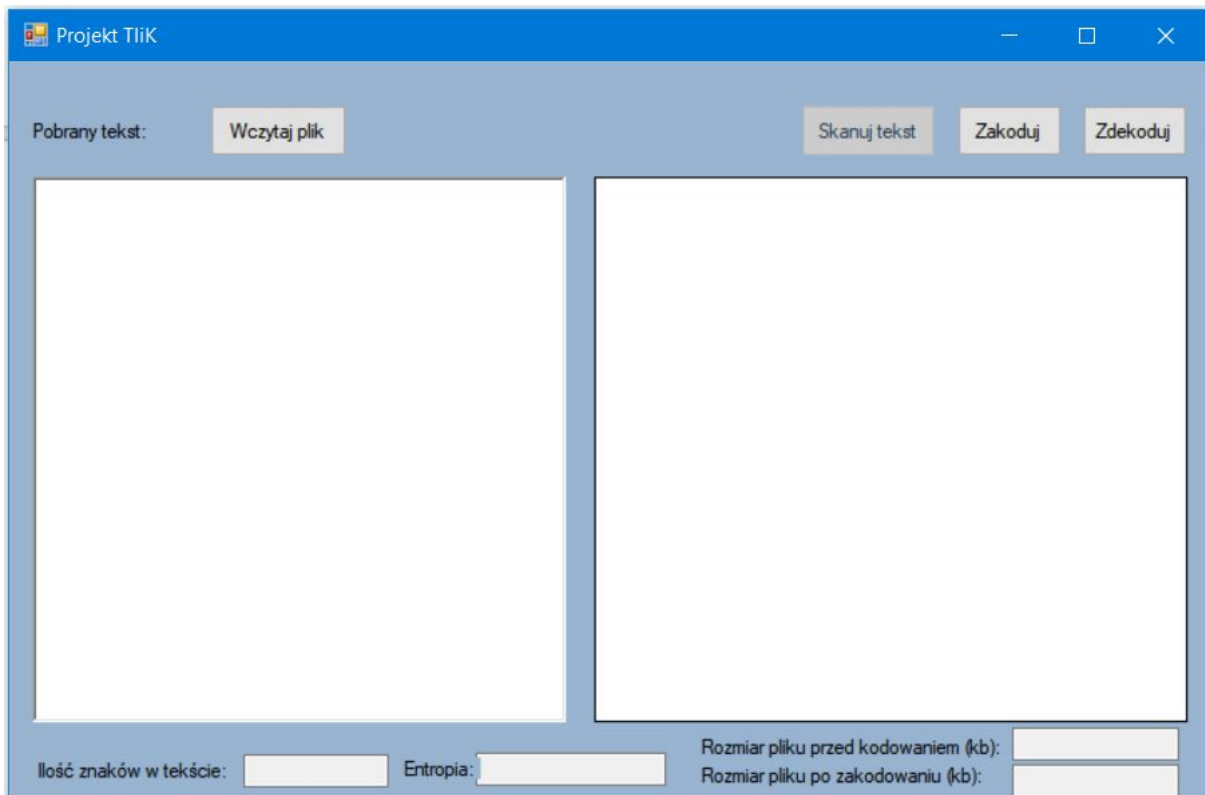
<b>Autor</b>	<b>Podzadanie</b>	<b>Zadania wspólne</b>
Damian Filipowicz	1. Kompresja pliku. 2. Przygotowanie plików testowych.	1. Dokumentacja.
Krzysztof Łuczak	1. Algorytm generowania słownika kodowego. 2. Zliczanie wystąpień par znaków.	
Maciej Marciniak	1. Interfejs graficzny. 2. Integracja wszystkich kodów.	
Dawid Wiktorski	1. Dekompresja pliku. 2. Obliczenie entropii.	

## IV. Realizacja

Rozdział prezentuje sposób realizacji projektu.

### 1. Implementacja graficzna

Pierwsze uruchomienie programu powoduje pojawienie się okna przedstawionego na Rys. 1. Możliwymi operacjami jakie można wykonać na “świeży” programie jest wczytanie tekstu z dysku oraz zdekodowanie pliku.



Rys. 1. Widok startowy programu

Przycisk “Wczytaj plik” otwiera okno z wyborem pliku z dysku, na którym planowane jest wykonywanie operacji. Zawartość tekstowa wyświetli się w białym polu po lewej stronie. Możliwe jest również wpisanie tekstu ręcznie. Po pojawieniu się tekstu w oknie odblokowany zostanie przycisk “Skanuj tekst”.

Wczytany tekst z użyciem funkcji “Skanuj tekst” skutkuje pojawieniem się po prawej stronie okna tabelki z statystykami wystąpień poszczególnych liter. Na dole zostanie uzupełnione pole związane z ilością znaków w pliku oraz entropią.

Wykonanie funkcji pod przyciskiem “Zakoduj” wykona w pierwszej kolejności analizę długości tekstu. Jeżeli tekst składa się z nieparzystej liczby znaków, dodawany na koniec jest znak spacji, wymagane jest to ze względu na charakter metody kodowania (pary znaków). Następnie tworzony jest słownik kodowy oraz wykonane statystyk tym razem dla pary znaków (program w języku Python). Wynik skryptu Pythonowego przekazany zostaje do funkcji kodującej wg algorytmu Shannona Fano pary znaków opisanego w punkcie I. Po zakończeniu procesu kodowania pojawia się okno, w którym należy wybrać lokalizację gdzie ma zostać zapisany zakodowany plik. Poprawnie zakończony proces kodowania powoduje wpisanie w polach związanych z rozmiarem pliku, wielkość przed i po kodowaniu.



## V. Wyniki kompresji

Tabela Tab 2. zawiera porównanie rozmiarów plików przed i po kompresji dla wszystkich plików testowych.

*Tab.2. Rozmiary plików przed i po kompresji*

Nazwa pliku	Rozmiar przed kompresją	Rozmiar po kompresji
HarryPotter.txt	25 kB	19 kB
raideng.txt	15 kB	12,4 kB
PanTadeusz.txt	13,1 kB	12,1 kB
raidpl.txt	19,1 kB	16 kB

## VI. Bibliografia

1. <https://www.youtube.com/watch?v=dJCck1OgsIA>
2. [http://lux.dmc.s.p.lodz.pl/ptda/wyklady/wyklad\\_3.pdf](http://lux.dmc.s.p.lodz.pl/ptda/wyklady/wyklad_3.pdf)
3. [https://pl.wikipedia.org/wiki/Kodowanie\\_Shannona-Fano](https://pl.wikipedia.org/wiki/Kodowanie_Shannona-Fano)
4. Dokumentacja biblioteki Collections języka Pythona,  
<https://docs.python.org/3/library/collections.html>