

MiscDev 001

Informacje wstępne i materiały dodatkowe

Informacje wstępne

W tej sekcji zawarte są informacje, z którymi uczestnik powinien zapoznać się przed przystąpieniem do oglądania videoartu.

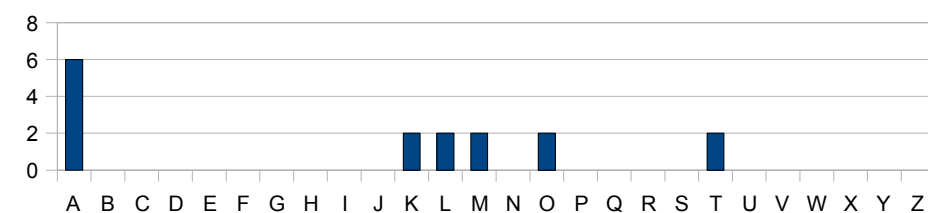
Histogram

W statystyce **histogram** jest wykresem (zazwyczaj w postaci szeregu prostokątów) liczebności występowania pewnej cechy w zbiorze.

W videoarcie histogram jest potraktowany nie tyle jako wykres, co jako tablica zawierająca liczebność występowania danej cechy w próbie empirycznej, a w szczególności, jako tablica w której każdy indeks odpowiada kolejnemu znakowi ASCII, a wartość elementów to liczebność danego znaku ASCII w próbie.

Przykładowo, dla ciągu „ALA MA KOTA KOT MA ALE” tablica taka wyglądała by następująco:

```
hist['A'] = 6
hist['B'] = 0
...
hist['K'] = 2
hist['L'] = 2
hist['M'] = 2
hist['N'] = 0
hist['O'] = 2
...
hist['T'] = 2
```



Rysunek 1: Fragment histogramu ciągu "ALA MA KOTA KOT MA ALE"

Histogram jako wykres oprawy o powyższą tablicę przedstawiony jest na rysunku 1. Dla uproszczenia histogram zawiera tylko wielkie litery, czyli znaki o kodach od 0x41 do 0x5A.

W kryptoanalizie, a konkretniej w kryptolingwistyce, istnieje metoda polegająca na odszyfrowywaniu wiadomości tekstowych, poprzez analizę częstotliwości występowania danych zakodowanych liter, oraz przyrównaniu ich częstotliwości do tablicy zawierającej częstotliwości występowania danej litery w języku w którym spodziewamy się że została napisana zaszyfrowana wiadomość. Czyli tworzy się histogram tekstu zaszyfrowanego, i porównuje do histogramu stworzonego z dużej próbki tekstu w danym języku (np. książki czy serii artykułów prasowych).

Dodatkowe materiały

Sekcja dodatkowe materiały zawiera zestaw linków oraz informacji rozszerzających zakres videoartu.

Linki

<http://www.open-std.org/> - strona zawierająca otwarte standardy, m.in. standard języka C oraz C++

<http://www.open-std.org/JTC1/SC22/WG14/www/docs/n1336.pdf> - najbardziej aktualny standard języka C w wersji C99; uczestnik powinien zapoznać się przynajmniej pobieżnie ze standardem, oraz nauczyć z niego korzystać

<http://www.asciitable.com/> - strona zawierająca bardzo dobrze przedstawioną tablicę kodów i znaków ASCII

<http://www.mingw.org/> - strona kompilatora MinGW GCC

Ćwiczenia i zadania

Zadanie 0

Przepisz z videoarta wszystkie finalne wersje napisanych programów, a następnie wyeliminuj literówki i przetestuj czy się kompilują i działają poprawnie.

Zadanie 1

Korzystając z programu hist stworzonego podczas videoarta, stwórz histogram dla języka polskiego oraz angielskiego. Wybierz w tym celu serię artykułów lub długi utwór literacki. Dla uproszczenia, wyeksportuj dany tekst do pliku tekstowego, oraz zignoruj polskie litery.

Zadanie 2

Zmodyfikuj program hist tak, aby wyrysowywał poziomy wykres na konsoli. Efekt powinien być mniej więcej następujący:

```
a ( 27) #####
b ( 1)
c ( 7) ###
d ( 5) ##
e ( 12) #####
f ( 7) ###
g ( 7) ###
h ( 5) ##
i ( 39) #####
j ( 0)
k ( 1)
l ( 5) ##
m ( 6) ##
n ( 18) #####
o ( 7) ###
p ( 8) ###
q ( 0)
r ( 13) #####
s ( 6) ##
t ( 22) #####
u ( 4) #
v ( 2)
w ( 2)
x ( 4) #
y ( 1)
z ( 5) ##
```

Podpowiedź: znajdź w tablicy maksymalną ilość wystąpień tej litery, uznaj że wykres w takim miejscu powinien być najwyższy (np. 50 znaków hash #), a następnie dla każdego elementu tablicy obliczaj $(\text{hist}[i]/\text{max}) * 50$ (przy czym pamiętaj, że tego typu operacje powinny być robione na liczbach zmiennoprzecinkowych). Wynikiem tego działania będzie ilość hashy które trzeba wypisać.

Zadanie 3

W programie do wyszukiwania stringów postaraj się wyeliminować wypisywanie stringów typu wide-char (dwa bajty na znak) które powstają w niepoprawnych miejscach, tj. np. w środku stringu ASCII. Najprostszą metodą zrobienia tego jest dodanie sprawdzania czy następnym znakiem po literze jest `\0` (kod ASCII 0) w przypadku stringów widechar.

Zadanie 4

Znając metodę wyłączenia zabezpieczenia regedit'a, zaproponuj 5 rzeczy które powinien zrobić programista regedit'a aby utrudnić obejście zabezpieczenia. Zastanów się również co mógłby zrobić programista systemu operacyjnego aby uniemożliwić skorzystanie z innych narzędzi do edycji rejestru.