KRY – Dokumentácia k projektu č.1

Vigenerova šifra

Ján Folenta (xfolen00), 4.4.2021

1 Implementácia

Implementácia obsahuje odhad dĺžky kľúča vstupného zašifrovaného textu pomocou Kasiskeho a Friedmanovho testu a samotný odhad kľúča. Vstupom programu je zašifrovaný text, ktorý je poslaný na štandardný vstup. Zo zašifrovaného textu sa pomocou funkcie *processCipher()* odstránia nežiadúce znaky, ako napríklad interpunkčné znamienka a medzery, a všetky písmená sa prevedú na malé písmená abecedy.

1.1 Friedmanov test

Friedmanov test používa index zhody, ktorý pomocou merania nerovnosti frekvencií písmen šifrovaného textu odhadne dĺžku kľúča. Index zhody je možné vypočítať pomocou vzťahu

$$I = \frac{\kappa_p - \kappa_r}{\kappa_o - \kappa_r}$$

kde κ_p je pravdepodobnosť, že 2 náhodne vybrané písmená abecedy sú rovnaké (v implementácii bola použitá hodnota 0,065 z prednášok), κ_r je pravdepodobnosť náhodne zvoleného písmena abecedy (zvolená bola hodnota 0,038 z prednášok) a κ_o je miera zhody, ktorú je možné vypočítať vzťahom

$$\kappa_o = \frac{\sum_{i=1}^{c} n_i(n_i - 1)}{N(N-1)}$$

pričom c je veľkosť abecedy, N je počet znakov šifrovaného textu a n_i sú frekvencie jednotlivých písmen.

Friedmanov test je implementovaný vo funkcii *friedmanTest(),* v ktorej sú spočítané zastúpenia jednotlivých písmen v šifrovanom texte a následne pomocou vzťahov popísaných vyššie je odhadnutá dĺžka kľúča.

1.2 Kasiskiho test

Kasiskiho test je implementovaný vo funkcii *kasiskiTest()*. Tento test spočíva v nájdení vzdialeností medzi rovnako zašifrovanými časťami šifrovaného textu. V implementácii Kasiskiho testu sú prehľadávané sú opakujúce sa trigramy. Trigramy, ktoré sú prehľadávané v šifrovanom texte sú z dôvodu šetrenia výpočetných zdrojov postupne vyberané iba z prvej polovice šifrovaného textu. V prípade nájdenia opakujúcich sa trigramov sú pomocou funkcie *findDistances()* ukladané vzdialenosti susedov. Následne funkcia *findDividors()* nájde delitele jednotlivých vzdialeností a uloží si ich. Výber odhadovanej dĺžky kľúča je realizovaný vo funkcii *mostFrequentDividor(),* kde sa vyberú najčastejšie sa opakujúce delitele, z ktorých je ako výsledná dĺžka vybraný najväčší deliteľ.

Keďže Kasiskiho test v porovnaní s Friedmanovym testom vykazoval lepšie výsledky a mal celkovo uspokojivú úspešnosť, tak odhadovaná dĺžka kľúča, ktorá je použitá na nájdenie kľúča je daná Kasiskiho testom.

1.3 Nájdenie kľúča

Nájdenie kľúča spočíva v postupnom dešifrovaní písmen kľúča a podobá sa na index koincidencie. Najprv rozdelíme šifru vo funkcii *findKey()* na segmenty podľa dĺžky kľúča. Pre každý znak kľúča tak dostaneme všetky písmená, ktoré boli zašifrované daným znakom. Následne je podľa odhadnutej dĺžky kľúča pre každý znak zavolaná funkcia *findLetter()*, ktorá nájde správny znak kľúča. Metóda je založená na frekvencii písmen anglickej abecedy, kde porovnávame frekvenciu písmen zašifrovaných jedným znakom s frekvenciou písmen typického anglického testu [1]. Zašifrované písmená po každom výpočte pomocou funkcie *shiftLeft()* posunieme smerom doľava (z písmena C sa stane B a podobne) a znova porovnáme podobnosť frekvencií. Tento postup je opakovaný 26 krát, keďže anglická abeceda má 26 znakov. Posun, ktorý je najbližšie frekvencii typického anglického textu hovorí o tom, o koľko pozícii napravo od písmena A sa nachádza správne písmeno kľúča. Pre výpočet podobnosti bol použitý vzťah

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - F_i)^2}{F_i}$$

kde f_i sú frekvencie zašifrovaných písmen a F_i sú frekvencie písmen typického anglického textu. Hodnota najbližšia nule znamená najväčšiu podobnosť. Takým spôsobom je odhadnutý každý znak kľúča, ktorý je potom vypísaný na štandardný výstup.

2 Použitie

Program je najprv nutné preložiť príkazom `make` a následne spustiť príkazom

```
./kry < input.txt
```

kde *input.txt* je textový súbor obsahujúci zašifrovaný anglický text. Výstupom programu je následne štvorica – výsledok Friedmanovho testu, výsledok Kasiskiho testu, stanovená dĺžka kľúča a samotný kľúč, pričom každý údaj je oddelený bodkočiarkou.

3 Použitá literatúra

[1] https://en.wikipedia.org/wiki/Letter_frequency