KRY 1. Projekt

Andrej Zaujec

April 4, 2021

Contents

1 Úvod

Cielom projektu je vykonanie Friedmanovho a Kasiskeho testu nad textom zašifrovaným Vigenerovou šifrou, následne zvoliť vhodnú dĺžku kľúča a taktiež zistiť použitý kľúč. Dokumentácia je rovnako logicky členená ako samotný kód. V jednotlivých kapitolách sú popísané zjednodušene jednotlivé moduly a taktiež rozhodnutia, ktoré boli počas implementácie vykonané.

2 Hlavný modul

Hlavný modul alebo taktiež nazývaný main, má za úlohu tieto kroky.

- Prečítať vstup
- Očistiť vstupný reťazec zmazaním všetkých znakov okrem znakov anglickej abecedy

- Spustiť Friedmanov test nad očistením reťazcom
- Spustiť Kasiskeho test nad očistením reťazcom
- Na základe výsledkov predchádzajúcich testov sa rozhodnúť pre správnu dĺžku klúča
- Uhádnuť klúč volaním triedy KeyFinder
- Vypísanie všetkých ziskaných hodnôt na štandartný výstup

Najzaujímavejšou činosťou tohoto modulu je zvolenie čo najsprávnejšieho kľúča, ktoré spočíva v nasledujúcich krokoch. Pogram sa pozrie na 10 najčastejších deliteľov získaných z Kasiskeho testu, získanie týchto deliteľov bude viacej popísané v kapitole 3. Následne tieto delitele považuje za možné dĺžky klúča. Každá vhodná dĺžka kĺúču musí byť maximálne dva krát tak veľká ako výsledok Friedmanovho testu. Toto rozhodnutie plynie z mojej skúsenosti počas testovania kedy pri neobmedzenej možnej dĺžke sa duplikovaná klúč na základe IC(Index of coincidence) javil ako vhodnejší kandidát než samotný správny, kratší klúč. Zároveň táto podmienka má aj svoje obmedzenie a to je prípad kedy je klúč príliš dlhý a Friedmanov test dva krát menší, v takomto prípade nebude správny kľuč nájdený. Zo spomínaných 10 kandidátov, ktorý sú kratší ako dvojnásobok Friedmanovho testu sa pre každý spočíta IC pre k podreťazcov, kde k je dĺžka klúča a následne sa tieto IC spriemerujú a kĺúč, ktorý má tento priemer najväčší je považovaný za najvhodnejšieho kandidáta na dĺžku klúča.

3 Kasiski modul

Ako už názov modulu naznačuje, tento modul je zameraný na výpočet Kasiskeho testu. Implementácia je na základe rozkladu vzdialeností opakujúcich sa ngramov na najčastejšie vyskytujúcich deliteľov. Tvorenie ngramov a počítanié ich výskytov a vzdialeností má na starosť modul Ngram popísaný v 4 kapitole. Dĺžky ngramov sú 4,5,6,7. Tieto veľkosti sa javili ako najoptimálnejšie počas testovania a trigramy som vylúčil pretože už boli viac krát zahrnuté v spomínaných dĺžkach ngramov a neprinášali žiadnu informáciu navyše. Výsledok Kasiskeho modulu je v mojom prípade zoznam dvojíc, kde prvý prvok dvojce je daný deliteľ a druhý prvok je počet jeho výskytov. Tento zoznam je zoradený podľa najčastejších výskytov deliteľov to znamená podľa druhého prvku. Ako celkový výsledok Kasiskeho testu volím najčastejšie vyskytujúceho deĺitela okrem čísla dva. Toto číslo som usúdil,

ako priveľmi krátke na to aby bolo kĺúčom a veľmi často sa objavovalo ako najčastejší delitel. Jedným z pozorovaných problémov tohoto prístupu je nie vždý vhodná voľba správneho klúča pretože väčšie kľuče sa častokrát predbehované svojími menšími deliteľmi a teda za výsledok sa zvolí len nejaký menší ďeliteľ spravného klúču.

4 Ngram modul

Tento modul slúži na výpočet ngramov zo zadaného textu a je používaný iba Kasiski modulom. Podstatou výpočtu ngramov je rozdelenie textu na ngramy o určitej dĺžke a následne sa frekvencia ngramov porovnáva s prievazaným zoznamov. Každý záznam v tomto previazanom zozname reprezentuje jeden ngram, počet výskytov daného ngramu a taktiež aj uchovanie vzdialeností rovnakého ngramu vo vstupnom texte. Previazaný zoznam je optimalizovaný, aby čo najčastejsie opakujúce ngramy radil na svoj začiatok.

5 Friedman modul

Daný modul je zameraný na výpočet Friedmanovho testu nad daným reťazcom. Výpočet klúča s dĺžkou k je daný vzorcom.

$$k = \frac{k_p - k_r}{k_0 - k_r}$$

$$k_0 = \frac{\sum_{i=1}^{c} n_i(n_i - 1)}{N(N-1)}$$

$$k_r = 0.0385$$

$$k_p = 0.067$$

Kde n_i je výskyt daného písmena v reťazci a N je veľkosť reťazca. Pričom konštanty k_p a k_r sú vhodne zvolené pre anglický monocase text. Ich dôsledný význam je možné nájsť na stránke zdroja samotného vzorca.

6 KeyFinder modul

Tento modul dokáže nájsť heslo pre daný reťazec a správne zvolenú dĺžku kĺúča k. Princíp hľadania klúča spočíva v rozdelení si reťazcu na k podreťazcov pričom znak na indexe n bude patriť do $n \pmod k$ podreťazcu pričom podreťazce označujeme od 0 po $k - 1$. Po tomto rozdelení už stačí

vyriešiť monoalfabetickú šifru nad každým podreťazcom pomocou nájdenie správneho posunu. Tento posun sa dá nájsť skrz vzorec.

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n}$$

Kde p_i je pravdepodobnosť písmena v danom jazyku, f_{i+g} sú frekvencie písmen anglickej abecedy v danom reťazci pričom g je spomínaný vhodný posuv. Hodnota M_g by sa v bežnom anglickom texte mala blížiť k 0.065. V tomto prípade spravím cyklus v cykle(skrz všetky posuvy a všetky písmená) a posuv, ktorú mi zaručil najbližšiu hodnotu M_g k 0.065 reprezentuje jedno písmeno kľúča anglickej monocase abecedy. Tento postup modul opakuje pre každý podreťazec a za predpokladu, že poznáme správnu dľžku klúča a zašifrovaný text je anglický a je používaný v bežnej reči tak sme získali klúč k správe zašifrovanej Vigenerovou šifrou. Zdroj vzorca a postupu je prednáška.