Криптографія

Лабораторна робота 3. Криптоаналіз афінної біграмної підстановки

ФБ-13 Ігнатенко Данило

Варіант 5

Мета роботи

Набуття навичок частотного аналізу на прикладі розкриття моноалфавітної підстановки; опанування прийомами роботи в модулярній арифметиці

Задача

0. Прочитати методичку. Тричі
1. Для статистики використовується повний текст із першої лаби (без літери "ъ")
2. Написати функції для пошуку оберненого елементу за модулем та розв'язків лінійних порівнянь
3. Знайти 5 найчастіших біграм шифротексту та мови (на основі тексту із першої лаби)
4. Перебором можливих співставлень частих біграм мови та шифротексту знайти кандидати на ключ
5. Перевірити змістовність розшифрованого ключем тексту
6. Знайти змістовний розшифрований текст та ключ(і) шифрування

Хід роботи

Цеглинка за цеглинкою: алгоритм Евкліда, на його основі – розв'язок порівнянь виду $ax = b \pmod m$, далі – шифрування та розшифрування текстів. Перевід числа у біграму відбувається просто за допомогою таблиці з усіма парами "число : біграма".

При розшифруванні перед записом відбувається перевірка тексту на змістовність на основі 2 критеріїв: заборонені біграми та частоти частих літер

Перший перевіряє наявність у тексті біграм з декартового добутку множин "аеийоуьыэюя" та "ьы". У російській мові такі буквосполучення не зустрічаються ні як частина слова, ні як остання те перша літера двох слів (звісно, є малий відсоток власних назв з інших мов, які, наприклад, починаються на літеру "ы", проте я дуже сумніваюся, що вони містяться у наданих нам текстах, а точніше я знаю, що всі надані тексти проходять за цим критерієм). Критерій відкидає приблизно 52% текстів*

Другий критерій порівнює відхилення частот літер "о", "а", "е" та "и" розшифрованого тексту від відповідних частот номінального тексту. Порогові значення відібрані наступним чином: з усіх 25 розшифрованих текстів знайдені найбільші відхилення для кожної літери, після чого їх збільшено на 10%. Текст бракується, якщо він не проходить хоча б за одним порівнянням. Коректно розшифрований текст з наданого набору гарантовано пройде перевірку за даним критерієм. При тестуванні критерій зменшив кількість текстів з решти ~24 тисяч до ~170*. Серед них – усі змістовні

* – під текстами маються на увазі файли, які виникали при розшифруванні. Усього таких очікувалося $C_{10}^2 \cdot C_{10}^2 \cdot 25 = 50625$. Вочевидь, серед них буде багато однакових, причому повторюються як змістовні, так і не змістовні. Також не було враховано відсутність варіантів розшифрування тексту 13, оскільки він потребує більше 10 найчастіших біграм. В кінцевому варіанті програма зупиняється при знаходженні одного змістовного тексту

Труднощі

Спершу було незрозуміло, як взагалі працює підстановка і як шифрувати та розшифровувати біграми, бо оперуємо ми з числовими представленнями

Далі потрібно було зрозуміти, за якими значеннями частот літер відкидати незмістовні тексти

Ну і ще треба було вигадати, як скоротити кількість циклів у функції massdec() з 4 до хоча б 2, бо спершу функція мала вигляд

```
for i in range(len(ptfq)-1):
    for j in range(i+1, len(ptfq)):
        for k in range(len(ctfq)-1):
            for l in range(k+1, len(ctfq)):
...
```

Шляхи розв'язання

Перша проблема вирішилася шляхом перебирання пари значень на папері, а числа в біграми переганяються за допомогою словника

Як були підібрані значення для оцінки змістовності було описано у Ході роботи

Остання проблема не така серйозна, але конструкцію для її вирішення я написати все ж примудрився

Результати

Повний відкритий текст збережений у файлі stat.txt (щоб порахувати частоти літер та біграм), зашифрований текст за варіантом – у файлі 05.txt, розшифрований – у файлі dec_05_(654, 777), де (654, 777) – ключ шифрування (для даного тексту – єдиний)

5 найчастіших біграм шифротексту – "вн", "тн", "дк", "хщ" та "ун"

```
(base) PS C:\Users\uranus\Desktop\Crypt\ihnatenko_fb-13_cp3> python lab3.py
Найчастіші біграми шифротексту
['вн', 'тн', 'дк', 'хщ', 'ун']

Найчастіші біграми мови (на основі лаб1)
['то', 'но', 'ст', 'ов', 'на']

Ключі розшифрування
a = 654, b = 777
```

Висновки

У ході роботи було реалізовано підпрограму для розв’язку лінійних порівнянь, яка використовується для розшифрування афінної біграмної підстановки на основі статистичних даних мови та шифротексту. Також було реалізовано простий розпізнавач російської мови на основі заборонений біграм та статистичних даних з наданих текстів