**Создание экспертной системы взлома полиалфавитного шифра (на примере шифра Виженера) с использованием частотного анализа, метода распределения вероятностей.**

**Этап 1. Естественное описание алгоритма**

*Описание Шифра Виженера*

*<…>*

**Состав алгоритма поиска ключа для взлома:**

1. Поиск длинны ключа;
   1. Поиск длинны методом индекса совпадений (У. Фридман);
   2. Организация выборки возможных длин ключа;
   3. Определение оптимальной длинны ключа;
2. Поиск ключевого слова;
   1. Определение вероятности различных сдвигов между группами методом частотного анализа;
   2. Организация выборки найденных вероятностей в виде таблицы (где строки – i-тый символа ключа, столбцы – символы алфавита);
   3. Организация ГА;
      1. Сопоставление полученных таблиц распределения вероятностей между собой (для выбранного языка, в данном случае Русский);
      2. Получение ключа на основе сопоставления вероятностей;
      3. Пробная дешифрация;
      4. Анализ текста;
      5. В случае неудачи модификация ключа;
      6. Переход к пункту *2.c.iii*
      7. При полной неудаче алгоритма возврат к поиску определения оптимальной длины ключа *1.c*.

**Этап 2. Описание алгоритма дешифровки.**

**Поиск длинны ключа**

Этот метод был предложен Уильямом Фридманом в 1922 году для взлома шифра Виженера. Метод основан на том факте, что вероятность совпадения двух случайных букв в некотором достаточно длинном тексте (индекс совпадений) — это постоянная величина. Таким образом, если разбить текст на L групп символов, каждая из которых зашифрована шифром простой замены (напомню, это и означает, что L — длина ключевого слова), то индексы совпадений для каждой из групп будут довольно близки к теоретическому значению этой величины; для всех других разбиений индексы совпадений будут гораздо ниже.

Рассмотрим текст, написанный на некотором языке. Алфавит данного языка будем полагать состоящим из **m** символов. Рассмотрим достаточно длинную строку **x** из **n** символов. Индексом совпадения называют вероятность совпадения двух произвольных символов в строке. Если **fi** — количество **i**-го символа алфавита в строке **x**, то индекс совпадения вычисляется по формуле:

Допустим, строка **x** является открытым текстом либо получена из него простой перестановкой. В этом случае индекс совпадений удобно выразить через вероятности появления **i**-го символа. Обозначим их **pi**. Тогда получим следующую формулу:

Т.к. величины **pi** имеют вполне определённые значения, то для открытого текста индекс совпадений не зависит от его содержания, а зависит только от языка, на котором написан текст. Более того, значения **pi** исследованы и известны, что позволяет рассчитать значения индекса совпадений открытого текста для различных языков.

| **Язык** | **Индекс совпадений** |
| --- | --- |
| **русский** | **0.0553** |
| английский | 0.0644 - 0.0667 |
| итальянский | 0.0738 |
| испанский | 0.0775 |
| немецкий | 0.0762 |
| французский | 0.0778 |
| ведийский санскрит | 0.021076696 |
| пракрит | 0.046635758 |
| классический санскрит | 0.045567736 |
| хинди | 0.041837864 |
| урду | 0.057535302 |

Далее мы разбираем получившиеся вероятности и ищем оптимальную для выбранного языка.

**Поиск ключевого слова**

Для получившейся длины ключа ищем вероятности для каждого i-ого символа ключа и организуем массив данных, где будут указаны вероятности сдвигов каждого символа ключа к символам алфавита.

**Сравниваем полученные результаты сопоставлением полученных вероятностей между собой.**

Рассмотрим две строки **x** и **y** с длинами **n** и **n'** соответственно. Алфавит, как и прежде, состоит из **m** символов. Взаимным индексом совпадений этих строк называют вероятность того, что символ, случайно выбранный из первой строки, совпадёт со случайно выбранным символом второй строки. Пусть **fi, gi** — количество **i-ого** символа алфавита в первой и второй строках соответственно. Тогда взаимный индекс совпадений будет равен:

Практически важным для метода индекса совпадений является частный случай, когда обе строки получены сдвигом алфавита открытого текста. Обозначим **pi** — вероятности появления **i-ого** символа в строке **x**, **s** — сдвиг алфавита строки **y** относительно алфавита строки **x** (влево). Тогда вероятности появление **i-ого** символа алфавита в строке **y** равны **pi + s** (используется нумерация алфавита строки **x**). Для взаимного индекса совпадений получаем следующую формулу:

Заметим, что т.к. сдвиг циклический, то

и взаимный индекс совпадений для сдвигов **s** и **m - s** принимает одно и то же значение.

Ниже приведены значения взаимного индекса совпадений в зависимости от сдвига для русского и английского языков. Значения приведены для сдвигов от **0** до **m/2**. Как упоминалось выше, на основе этих значений взаимный индекс совпадений может быть вычислен для любого сдвига.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Для русского языка:   |  |  | | --- | --- | | **Сдвиг** | **Взаимный индекс** | | 0 | 0,0553 | | 1 | 0,0366 | | 2 | 0,0345 | | 3 | 0,0400 | | 4 | 0,0340 | | 5 | 0,0360 | | 6 | 0,0326 | | 7 | 0,0241 | | 8 | 0,0287 | | 9 | 0,0317 | | 10 | 0,0265 | | 11 | 0,0251 | | 12 | 0,0244 | | 13 | 0,0291 | | 14 | 0,0322 | | 15 | 0,0244 | | 16 | 0,0249 | | Для английского языка:   |  |  | | --- | --- | | **Сдвиг** | **Взаимный индекс** | | 0 | 0,0644 | | 1 | 0,0394 | | 2 | 0,0319 | | 3 | 0,0345 | | 4 | 0,0436 | | 5 | 0,0332 | | 6 | 0,0363 | | 7 | 0,0389 | | 8 | 0,0338 | | 9 | 0,0342 | | 10 | 0,0378 | | 11 | 0,0440 | | 12 | 0,0387 | | 13 | 0,0428 | |

Список литературы

[Генетический алгоритм. Просто о сложном / Хабр (habr.com)](https://habr.com/ru/post/128704/)

[Индекс совпадений — Википедия (wikipedia.org)](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D0%B4%D0%B5%D0%BA%D1%81_%D1%81%D0%BE%D0%B2%D0%BF%D0%B0%D0%B4%D0%B5%D0%BD%D0%B8%D0%B9)

[Взлом шифра Виженера с помощью частотного криптоанализа / Хабр (habr.com)](https://habr.com/ru/post/221485/)

[НОУ ИНТУИТ | Лекция | Классические шифры (intuit.ru)](https://intuit.ru/studies/courses/13837/1234/lecture/31196?page=7)

[Online Python - IDE, Editor, Compiler, Interpreter (online-python.com)](https://www.online-python.com/)

[Advertising as persuasion (catchenglish.ru)](https://catchenglish.ru/teksty/teksty-dlya-9-go-klassa/advertising-as-persuasion.html)

[Онлайн-конвертер текста в нижний регистр — MrTranslate.ru](http://mrtranslate.ru/tools/tolower.html)

[Удалить ненужные символы - TextTools.ru](https://texttools.ru/remove-unwanted-characters)

[Modular conversion, encoding and encryption online — Cryptii](https://cryptii.com/)