

# ТЕОРЕМА О БЕСКОНЕЧНЫХ ОБЕЗЬЯНАХ И АНАЛИЗ ТЕКСТА НА ЕЁ ОСНОВЕ

Рассматривается обоснование и попытки доказательства теоремы. А также методы анализа текста написанного "обезьянами".

## ВВЕДЕНИЕ

Теорема о бесконечных обезьянах (в одном из многочисленных вариантов формулировки) утверждает, что абстрактная обезьяна, ударяя случайным образом по клавишам пишущей машинки в течение неограниченно долгого времени, рано или поздно напечатает любой наперёд заданный текст. Если перенести данные рассуждения в обозримый масштаб, то теорема будет утверждать, что если в течение продолжительного времени случайным образом стучать по клавиатуре, то среди набираемого текста будут возникать осмысленные слова, словосочетания и даже предложения.

## I. ОБОСНОВАНИЕ ТЕОРЕМЫ

Согласно теореме об умножении вероятностей, если два события статистически независимы, то есть результат одного события не влияет на результат другого, то вероятность наступления обоих событий вместе равняется произведению вероятностей этих событий [2]. Теперь предположим, что пишущая машинка имеет 50 клавиш, а слово, которое должно быть напечатано – «банан». Если ударять по клавишам случайным образом, вероятность того, что первым напечатанным символом будет буква «б», равна  $1/50$ ; такова же вероятность того, что вторым напечатанным символом будет «а», и так далее. Эти события независимы; таким образом, вероятность того, что первые пять букв составят слово «банан», равна  $(1/50)^5$ . По той же причине вероятность того, что следующие 5 букв снова окажутся словом «банан», также равняется  $(1/50)^5$ , и так далее. Несложно вычислить вероятность того, что блок из 5 случайным образом напечатанных букв не окажется словом «банан». Она равна  $1 - (1/50)^5$ . Поскольку каждый блок печатается независимо, вероятность того, что ни один из первых  $n$  блоков по 5 букв не совпадает со

словом «банан», равна:

$$P = \left(1 - \frac{1}{50^5}\right)^n. \quad (1)$$

При увеличении  $n$ , как видно из формулы,  $P$  уменьшается.

Таблица 1 –

n	P(в процентах)
1000	99,999
1000000	99,68
100000000	73
1000000000	4

## II. АЛГОРИТМЫ ПОИСКА В СТРОКЕ

Постановка задачи: Пусть задан массив  $T$  из  $N$  элементов и массив  $W$  из  $M$  элементов, причем  $0 < M \leq N$ . Поиск строки обнаруживает первое вхождение  $W$  в  $T$ , результатом будем считать индекс  $i$ , указывающий на первое с начала строки (с начала массива  $T$ ) совпадение с образцом (словом). Были использованы три алгоритма поиска:

1. Алгоритм прямого поиска;
2. Алгоритм Д. Кнута, Д. Мориса и В. Пратта (КМП-поиск).
3. Алгоритм Р. Боуера и Д. Мура (БМ-поиск);

## III. ВЫВОДЫ

Из трёх используемых алгоритмов наилучшим оказался Алгоритм Р. Боуера и Д. Мура (БМ-поиск). Для нахождения результат он потратил наименьшее количество времени как и наименьшее количество итераций.

## Список литературы

1. <https://habrahabr.ru/>
2. <http://ru.math.wikia.com/wiki/>
3. <https://ru.wikipedia.org/wiki>

*Кондратюк Глеб Игоревич*, студент 2 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, kondratukgleb76@gmail.com.

*Кулыба Вадим Александрович*, студент 2 курса факультета информационных технологий и управления Белорусского государственного университета информатики и радиоэлектроники, kulyba.vadim@mail.ru.

*Научный руководитель: Гуринович Алевтина Борисовна*, кандидат физико-математических наук, доцент, заместитель декана, gurinovitch@bsuir.by.