**СОДЕРЖАНИЕ**

# Введение

**Актуальность**: поиск подстроки является одной из ключевых задач в компьютерных науках, находящий применение в обработке текстов, системах антиплагиата и других областях. Эффективность алгоритмов поиска подстроки, зачастую, определяет производительность таких систем.

**Цель работы**: разработать программную реализацию алгоритма Кнута-Морриса-Пратта(КМП) для поиска подстроки и интегрировать его в веб-сервис для практического использования.

**Объект исследования**: алгоритм Кнута-Морриса-Пратт, предназначенный для поиска подстроки в тексте.

Предмет исследования: реализация префикс-функции, которая лежит в основе алгоритма КМП

**Задачи**:

1. Провести теоретический анализ задачи поиска подстроки и методов ее решения.
2. Сравнить точность и скорость работы КМП и других алгоритмов поиска подстроки.
3. Реализовать алгоритм КМП на языках программирования Python и C++ с подробными комментариями
4. Выполнить юнит-тестирование алгоритма
5. Создать веб-интерфейс на JavaScript для взаимодействия с пользователем

**Практическая значимость**: Результаты исследования могут быть использованы в системах антиплагиата для быстрого и точного выявления заимствований в текстах.

Методы исследования:

Структура работы:

# Теоретический анализ задачи и алгоритмов её решения

## Формальное описание выбранной задачи

**Наименование**: задача поиска подстроки

**Математическая модель**:

* Дано две строки: текст (text) длиной (n) и подстрока (p) длиной (m), при этом 1 **≤** m **≤** n
* Цель: найти все индексы (i), начиная с которых подстрока (p) содержится в тексте(text), если такого индекса нет – вернуть отрицательный индекс
* Целевая функция: минимизация времени исполнения алгоритма, выраженная через асимптотическую сложность.

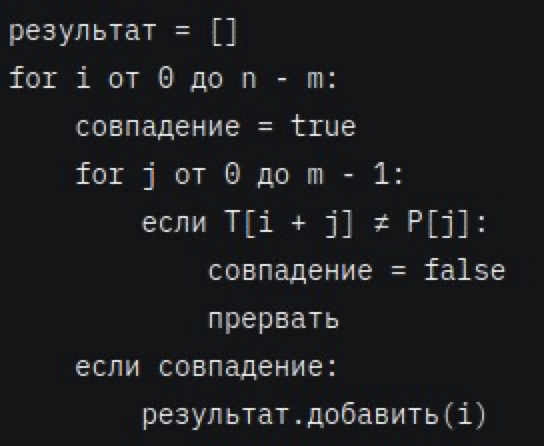
## Исследование алгоритмов решения задачи

Рассмотрим алгоритмы решения задачи поиска подстроки, напишем псевдокод, разберем принцип работы.

### Примитивный алгоритм

Идея данного алгоритма заключается в проверке каждой возможной позиции i в тексте T и сравнении символов с подстрокой P, если совпадение найдено, то индекс записываем результат.

С псевдокодом данного алгоритма можно ознакомиться на рисунке 1



Асимптотическая сложность данного алгоритма составляет O(n\*m), так как подстрока длинной m перебирается n раз.

### Алгоритм Кнута-Мориса-Пратта(КМП)