

# **КДЗ-2. Практический анализ алгоритмов поиска вхождений строки-шаблона в тексте**

**Проектирование и анализ алгоритмов**

**2022-2023 учебный год**

# Цель работы

- Реализация различных алгоритмов поиска вхождений заданного паттерна (шаблона) в исходном тексте
- Экспериментальный анализ временной сложности алгоритмов
- Сопоставление теоретических и практических оценок временной сложности алгоритмов
- Интерпретация экспериментальных результатов

# Рассматриваемые алгоритмы

- Наивный алгоритм с использованием простого посимвольного сравнения
- Алгоритм Кнута-Морриса-Прата с применением стандартных граней
- Алгоритм Кнута-Морриса-Прата с применением уточненных граней
- ???

# Подготовка тестовых данных. Часть 1

## Шаблоны без символа подстановки «?»

### Текст для поиска шаблона

1. Случайно сгенерированный текст в бинарном алфавите
  - 10000 символов
  - 100000 символов
2. Случайно сгенерированный текст в алфавите из 4 символов (а-ля последовательность ДНК)
  - 10000 символов
  - 100000 символов

### Искомый шаблон

Фрагмент исходного текста с некоторой позиции  $i$  переменной длины:

- 100-3000 символов
- Шаг длины шаблона - 100

# Подготовка тестовых данных. Часть 2

## Использование символа подстановки «?» в шаблонах

Текст для поиска используется тот же, что был получен ранее по схеме генерации предыдущего слайда.

В шаблон встраивается символ подстановка, который может означать любой символ из исходного алфавита.

Например, шаблон с двумя символами подстановки «a?dc?» входит в строку «abdcaddcd» дважды.

### Искомый шаблон

Фрагмент исходного текста с некоторой позиции  $i$  переменной длины:

- 100-3000 символов
- Шаг длины шаблона - 100

В случайную позицию полученного шаблона встраивается от 1 до 4 символов подстановки.

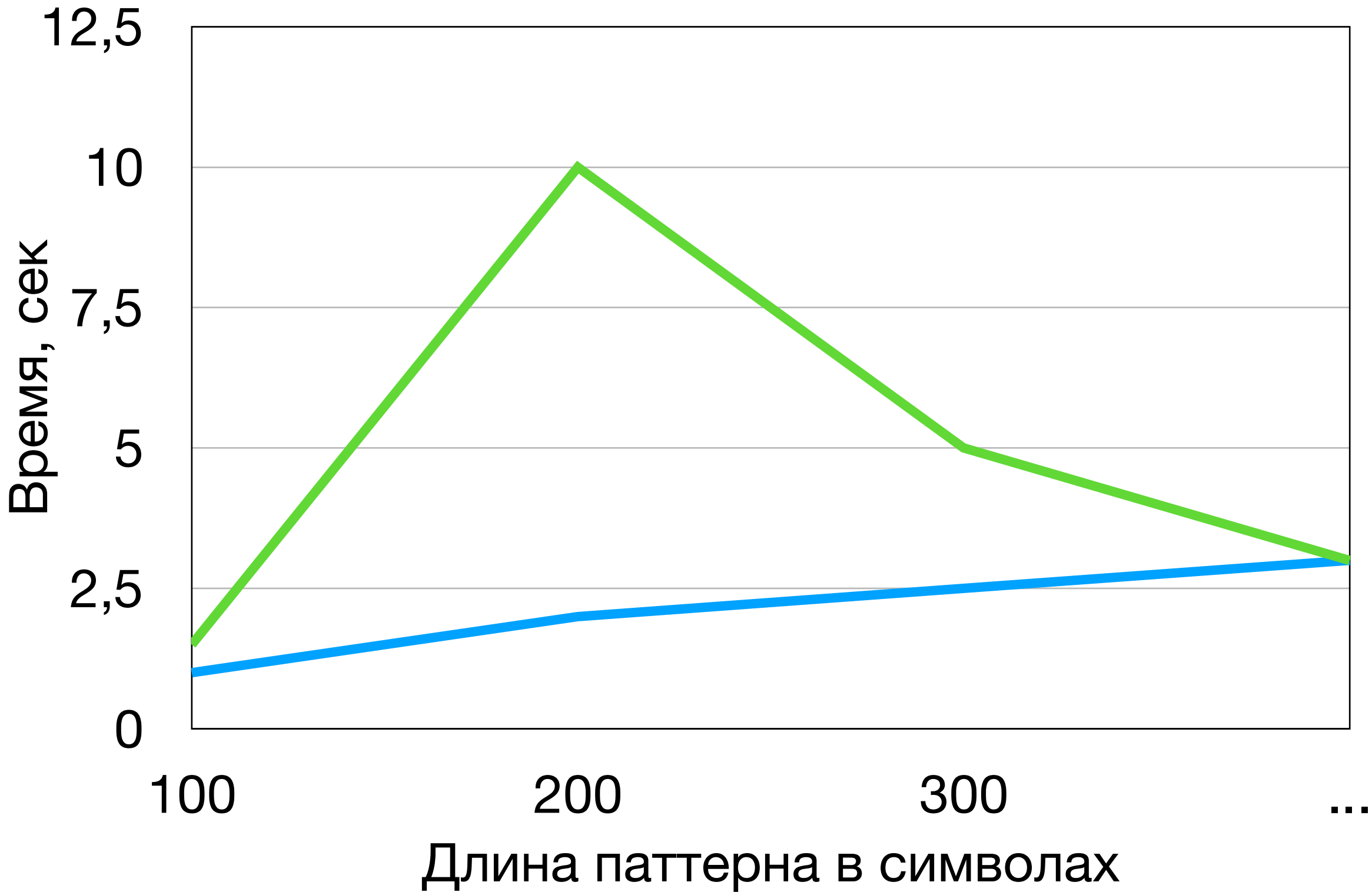
# Выходные сравнительные графики

- По всем алгоритмам:
  - Без символов подстановки, 10000 символов, бинарный алфавит
  - Без символов подстановки, 10000 символов, 4-символьный алфавит
  - Без символов подстановки, 100000 символов, бинарный алфавит
  - Без символов подстановки, 100000 символов, 4-символьный алфавит
  - С X символов подстановки, 10000 символов, бинарный алфавит
  - С X символов подстановки, 10000 символов, 4-символьный алфавит
  - С X символов подстановки, 100000 символов, бинарный алфавит
  - С X символов подстановки, 100000 символов, 4-символьный алфавит

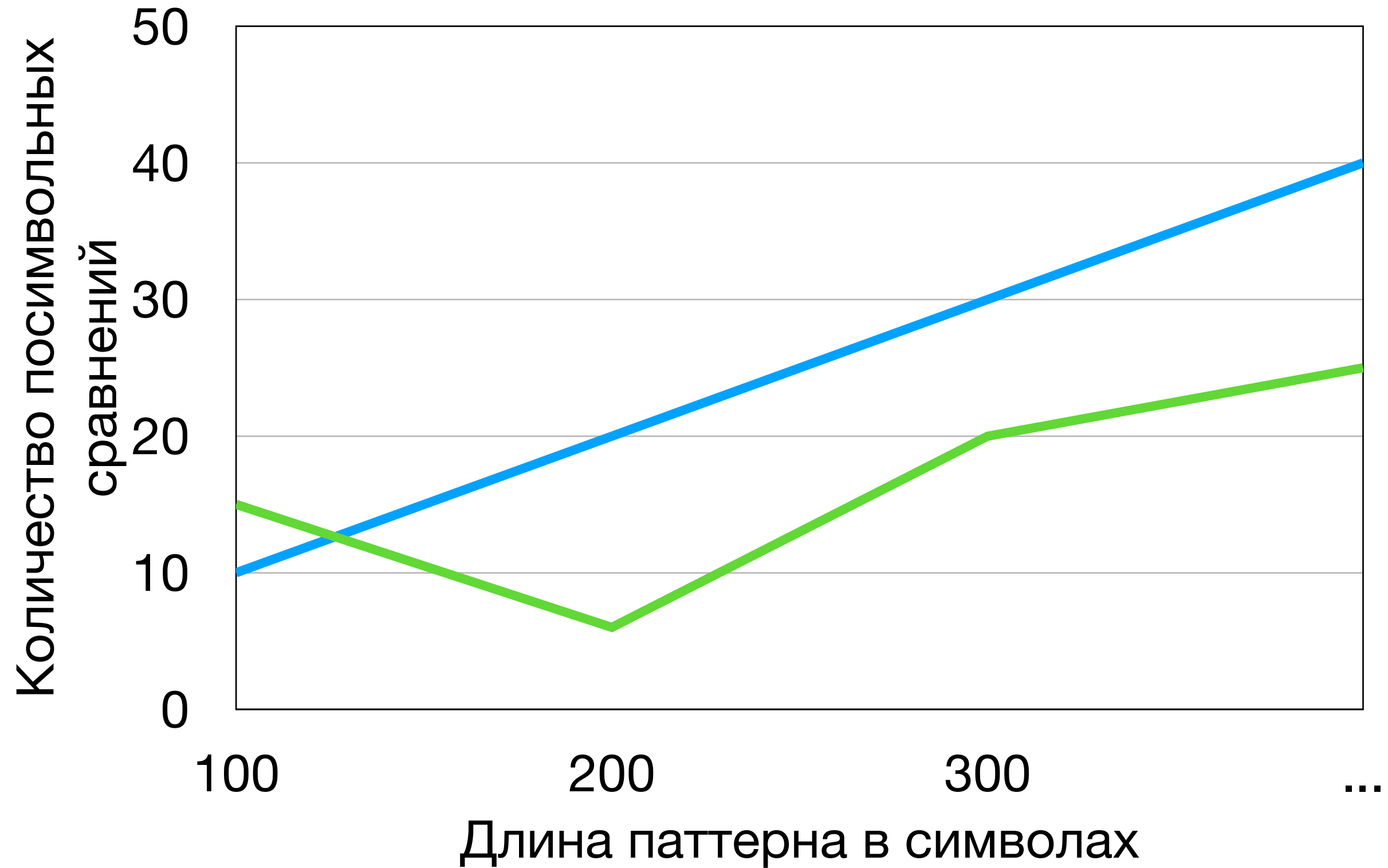
# Выходные сравнительные графики (примерный вид)

## Спецификация графиков

0 символов подстановки, 10000 символов,  
бинарный алфавит



0 символов подстановки, 10000 символов,  
бинарный алфавит



# Выходные сравнительные графики

## Дополнительные замечания

- Усреднение результатов замеров можно проводить любым удобным способом:
  - Большое количество тестирований
  - Медианные значения
  - и проч.



# Комплект выходной документации

- Исходный код
- Сравнительные графики
- Краткий текстовый отчет о проделанной работе с выводами о сравнении алгоритмов в соответствии с полученными результатами
- Иное, что **необходимо**