

# Выравнивание двух строк по алгоритму Нидлмана-Вунша. Отчет по задаче.

Владимирова Элина

21 декабря 2020 г.

## 1 Резюме

Для решения задачи об оптимальном выравнивании двух последовательностей был реализован классический алгоритм Нидлмана-Вунша.

Входные данные - длины двух строк. Преобразованиям подвергаются построенные по ним путем случайной генерации символов из заданного заранее алфавита буквенные последовательности. Результат работы программы - две строки, полученные из исходных вставками пропусков для достижения наибольшей посимвольной схожести последовательностей при их наименьшей длине.

## 2 Постановка задачи

Необходимо найти оптимальное выравнивание 2 заданных последовательностей, состоящих из символов алфавита  $\{A, G, C, T, \}$  т.е. выявить наименее трудозатратную комбинацию элементарных преобразований строк, позволяющую получить последовательности с максимальным совпадением символов на одинаковых местах, имеющие при этом минимальную длину.

## 3 Допущения

Строки генерируются внутри программы, их длины задаются вручную. Алгоритм, однако, не привязан к содержимому строк, что позволяет регулировать как алфавит, так и непосредственно строки.

## 4 Описание решения

Был реализован классический алгоритм Нидлмана-Вунша, предусматривающий работу с матрицей размера  $(n + 2) * (m + 2)$ , где  $n, m$  - длины строк.

## 5 Результаты

Правильность решения была проверена на строках малой длины (порядка 15 символов), в том числе на строках с сравнительно большим относительно длин модулем разности длин. Программа

успешно обрабатывает строки длиной до 12 тыс. знаков - по преодолении этого порога возникают проблемы с памятью, причиной которым - невозможность хранения выстраиваемой в ходе решения матрицы и произведения внутри нее предусмотренных алгоритмом операций.