|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Теоретическая информатика и компьютерные технологии

**Отчет по лабораторной работе № 6**

***по курсу «FASTA»***

**«Алгоритм Нидлмана-Вунша»**

Студент ИУ9-71Б **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**  Максимов Е.Н.

(Группа) (Подпись, дата) (Фамилия И.О.)

Преподаватель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Страшнов П. В. (Подпись, дата) (Фамилия И.О.)

*Москва, 2022 г.*

# Цель работы

Реализовать алгоритм FASTA.

# Постановка задачи

Для заданной последовательности нуклеотидов выполнить поиск в соответствующей базе данных последовательностей, чтобы найти совпадения с аналогичными последовательностями базы данных.

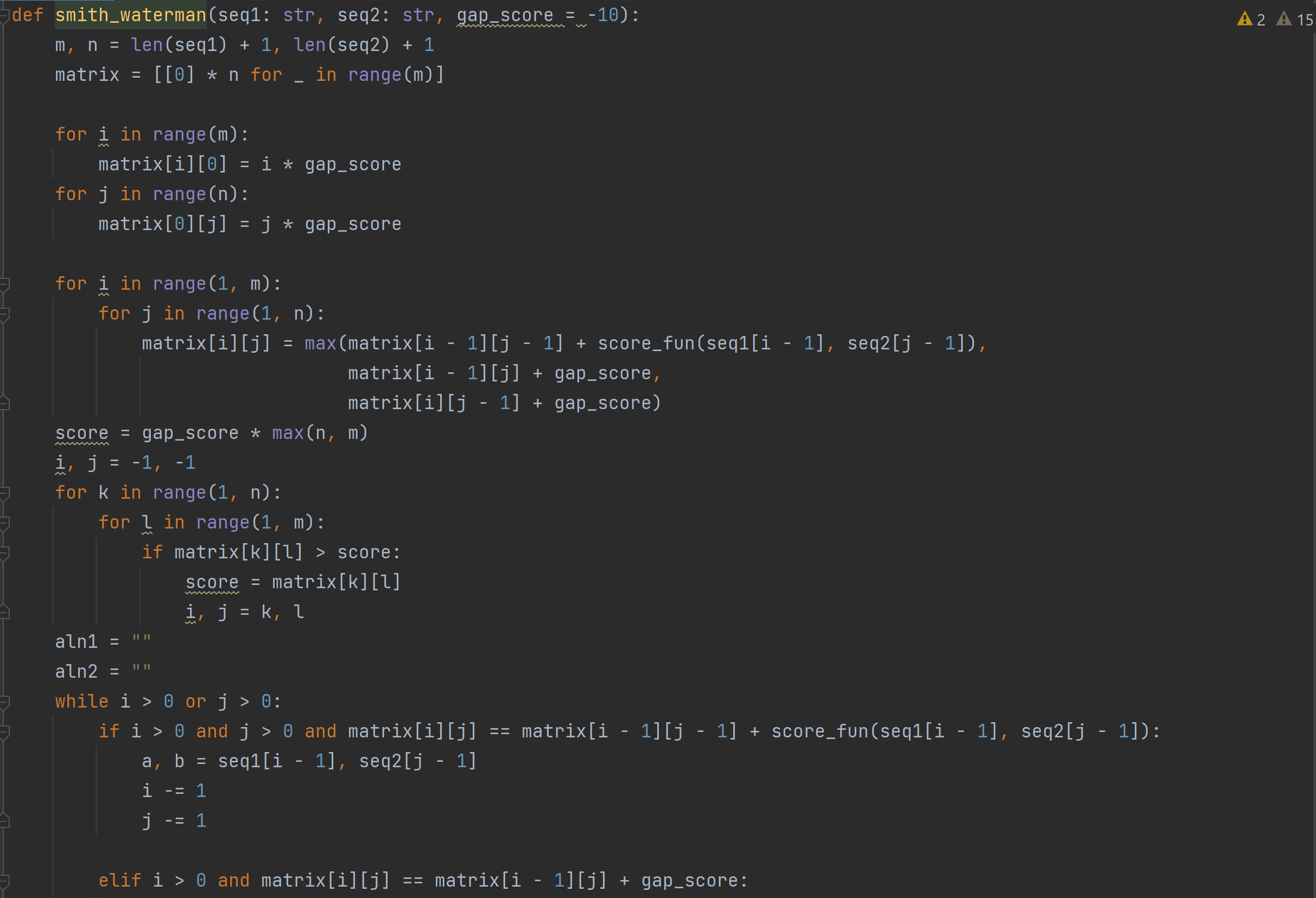
# Теоретическая часть

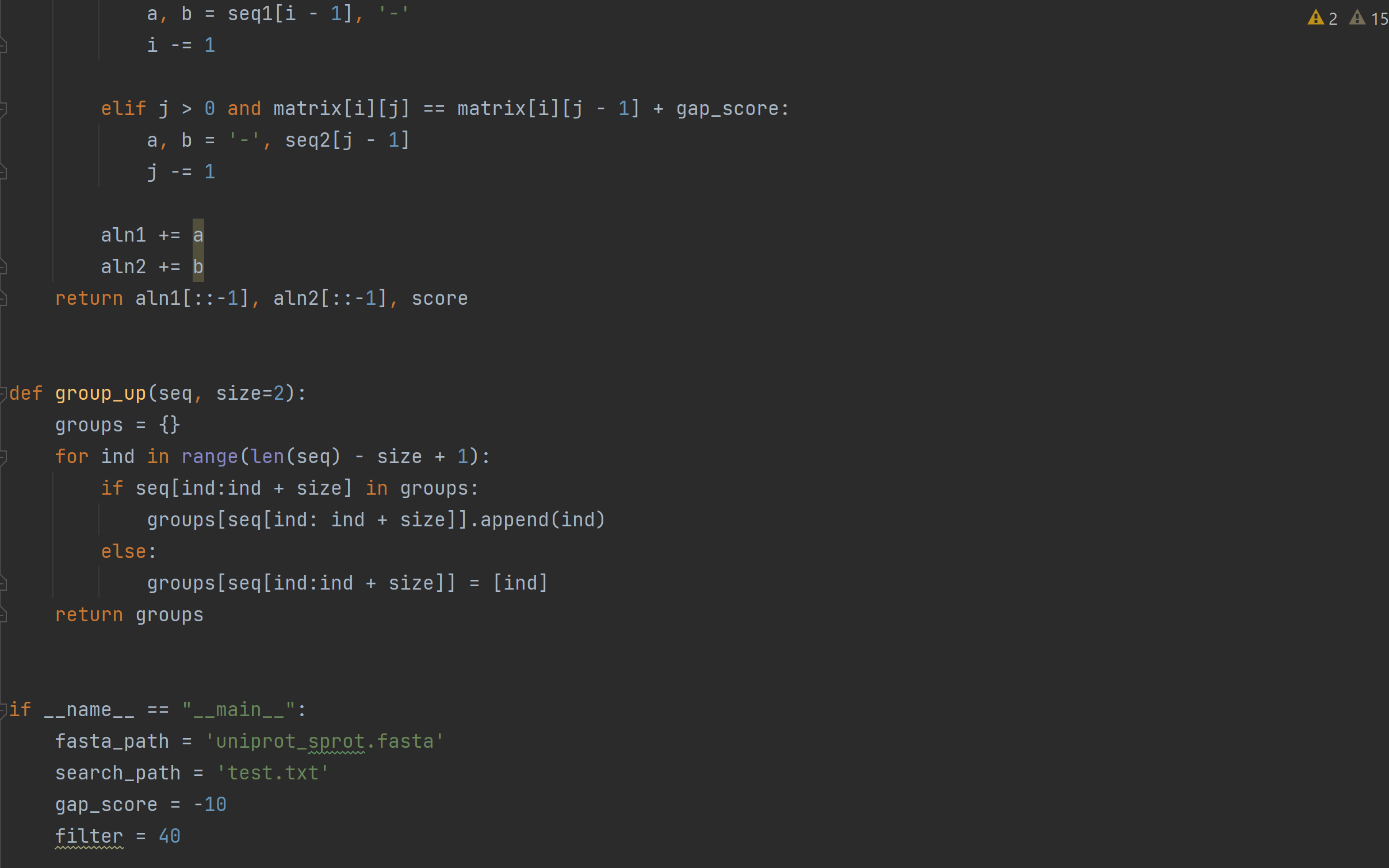
1. Сначала определяются области с наибольшей плотностью в каждом сравнении последовательностей. Последовательности разбиваются на пары нуклеотидов и затем осуществляется все сходные локальные области, представленные в виде диагоналей определенной длины на точечной диаграмме, между двумя последовательностями, подсчитывая совпадения и штрафуя промежуточные несовпадения. Таким образом, локальные области совпадений с наибольшей плотностью по диагонали изолируются от фоновых попаданий.
2. После этого производим отсеивание тех участков, которые не соответствуют заданному скору, вычисляется оптимальное выравнивание начальных регионов, как комбинация совместимых регионов с максимальным количеством баллов.
3. Используя групповой алгоритм Смита-Уотермана, производится вычисление оптимальной оценки для выравнивания. Для каждого выравнивания последовательности с последовательностью базы данных.

# Реализация

В качестве языка реализации был выбран Python:

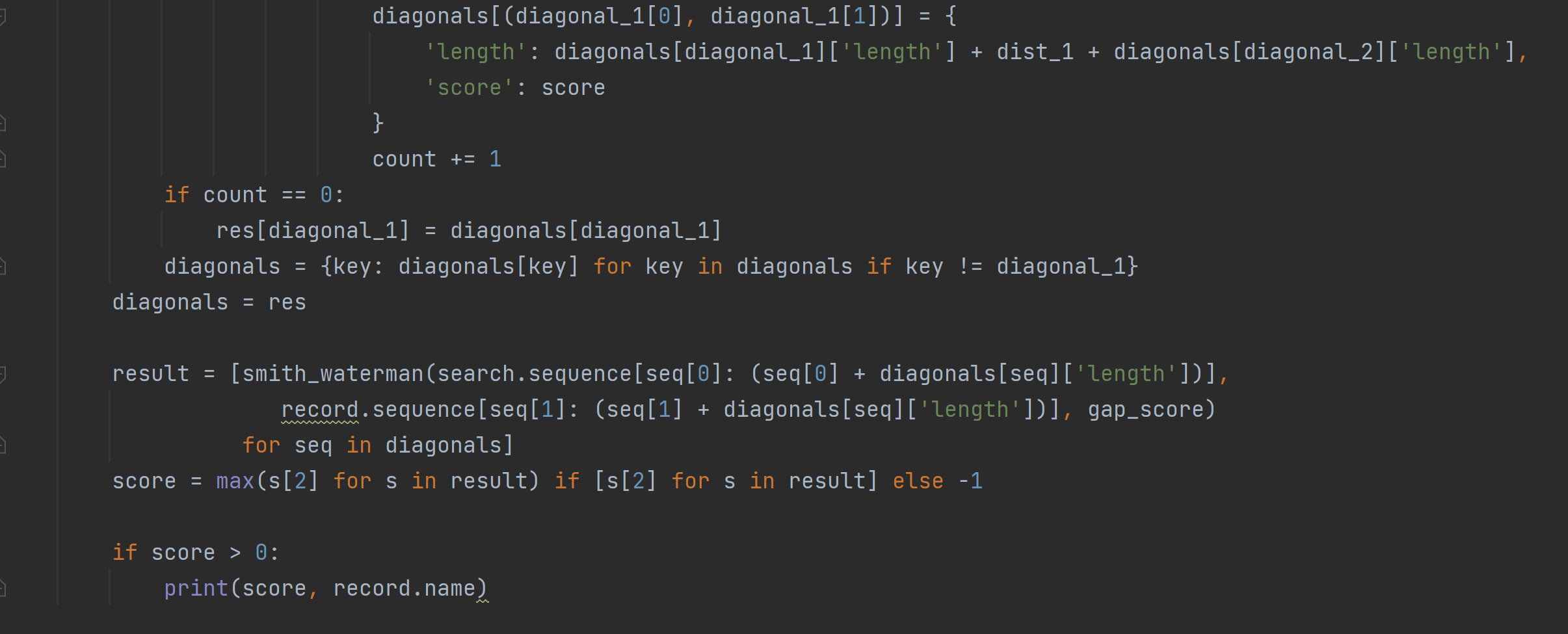








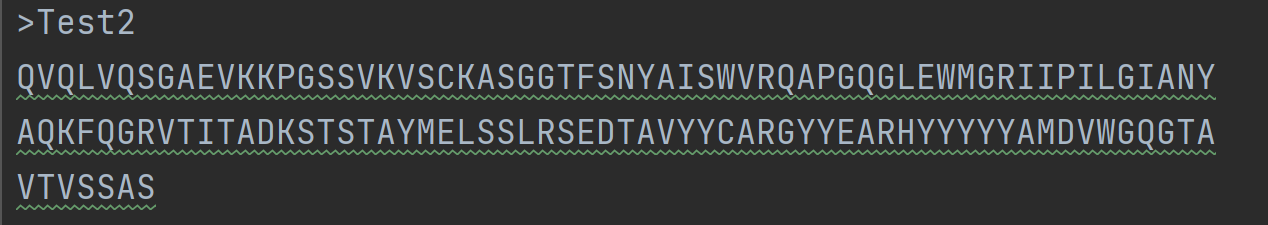


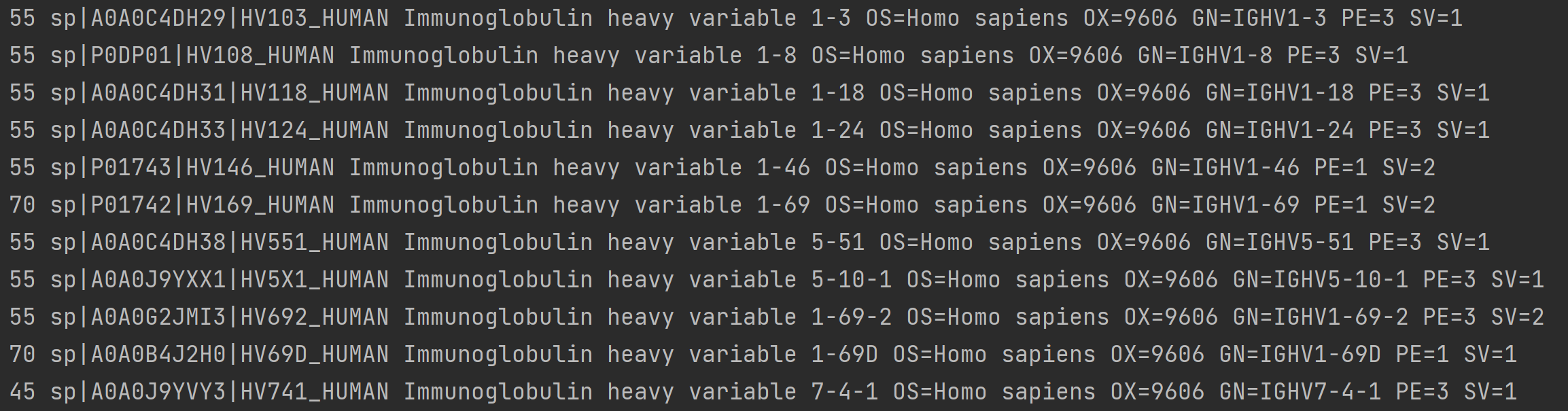


# Тестирование

Для воспроизведения тестирования следует скачать файл uniprot\_sprot.fasta, который используется как база данных и из файл test.txt берутся тесты. Для изменения базы данных и тестового файла следует поменять значения в строках 82, 83.

Тестовые данные:



Результат:

Проверка первого найденного с помощью внешнего ресурса:

