|  |  |
| --- | --- |
|  | **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 № 5**

***по курсу «Алгоритмы биоинфоматики»***

**«Алгоритм UPGMA»**

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 UPGMA.

# Постановка задачи

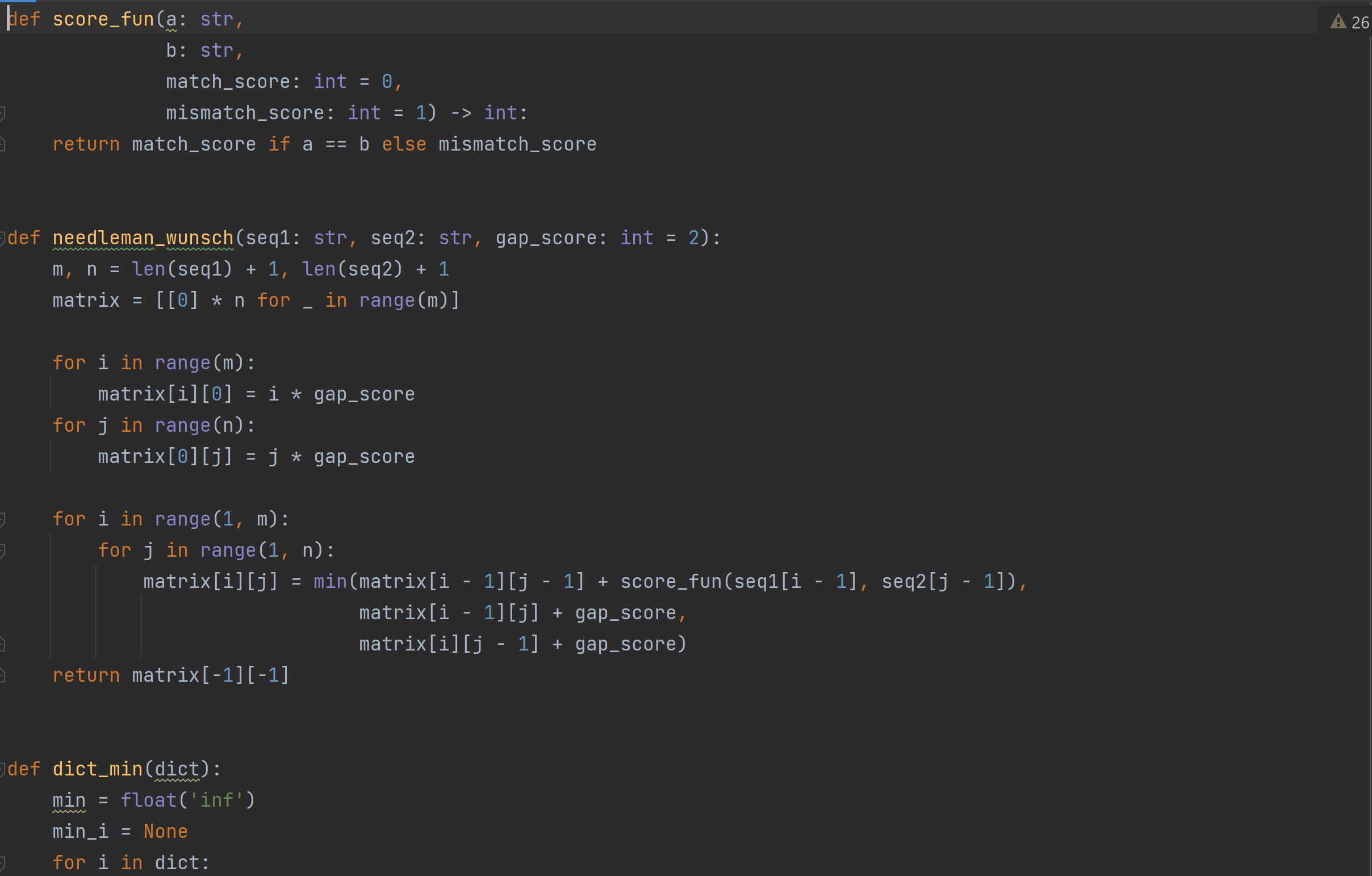
Выполнить построение филогенетического дерева объектов.

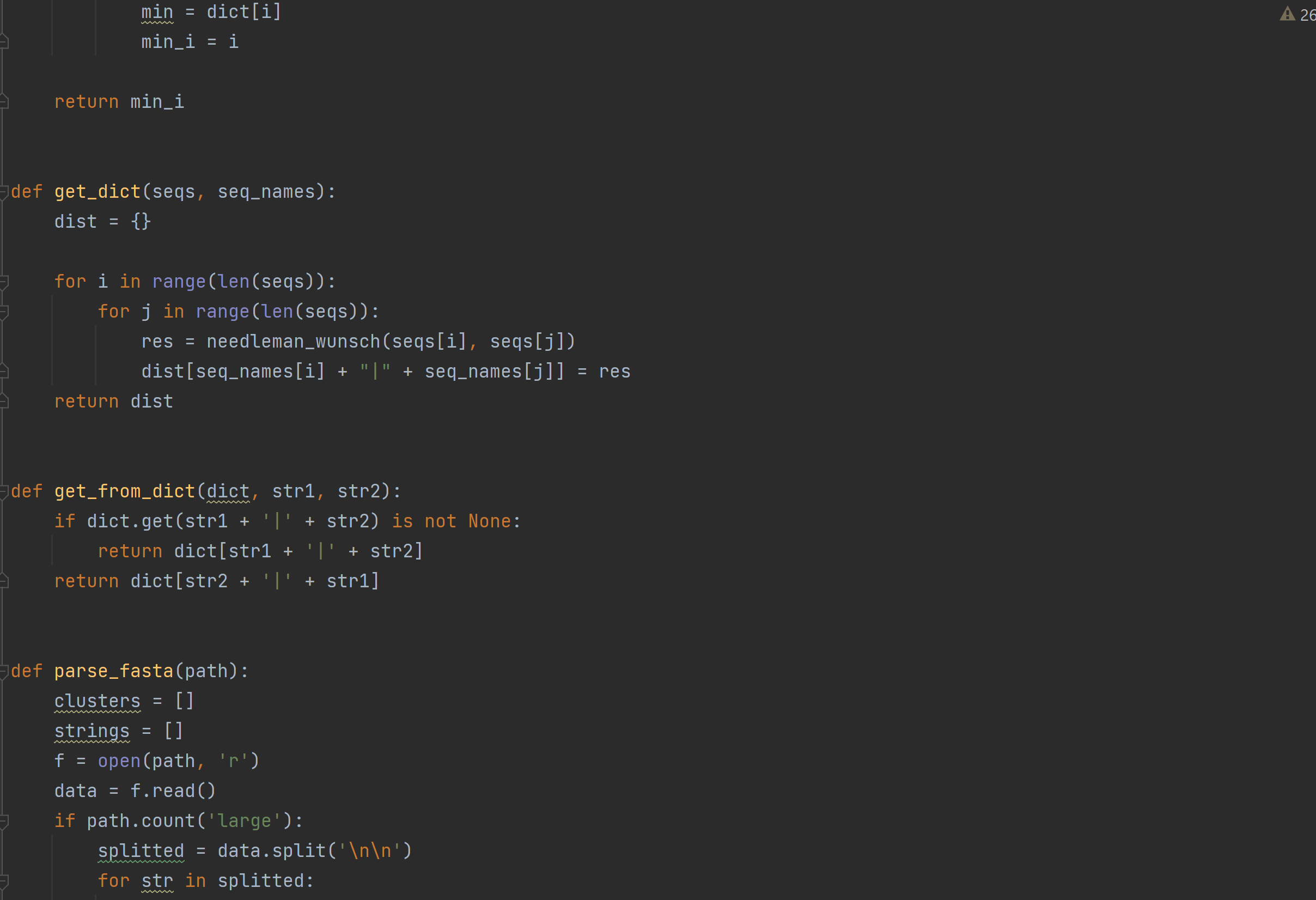
# Теоретическая часть

1. Выбираем наименьшее число в матрице расстояний. Если сразу несколько чисел имеют одинаковое (и меньшее остальных) значение, выбираем одно из них случайным образом. Запоминаем пару листьев, расстояние между которыми минимально (назовем их A и B).
2. Создаем новый узел дерева. Соединяем его ребрами с выбранными листьями A и B.
3. Приписываем вновь созданным ребрам одинаковые длины, равные половине значения расстояния между A и B (взятого из входной матрицы расстояний).
4. Создаем новую матрицу расстояний. Объекты, расстояния между которыми описаны матрицей, теперь двоякого рода: это, во первых, все листья, кроме A и B, а во-вторых, множество ("кластер") {A, B}. Расстояния между листьями остаются теми же, что были во входной матрице, а расстояние от произвольного листа X до кластера {A, B} считается по формуле: d(X, {A,B}) = (d(X,A)+d(X,B))/2, где расстояния d(X,A) и d(X,B) берутся из входной матрицы (то есть расстояние до кластера полагаем равным среднему арифметическому двух расстояний до его элементов).

# Реализация

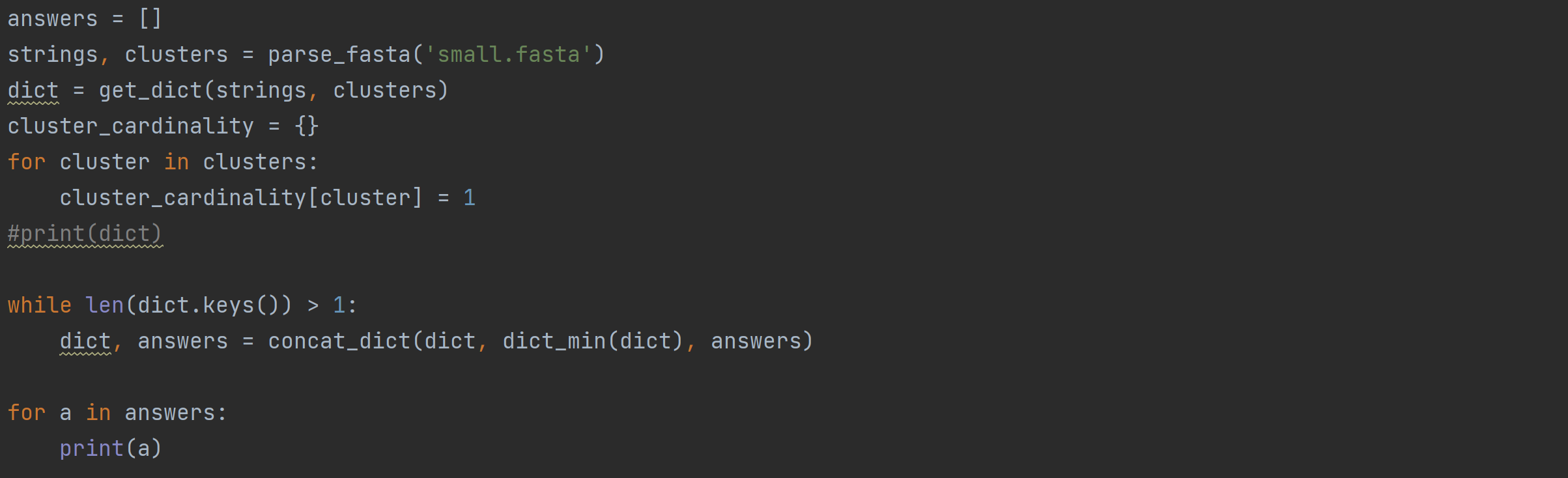
В качестве языка реализации был выбран Python:











# Тестирование

Для тестирования используется источник .fasta, воспользоваться которым можно добавив его в функцию parse\_fasta, вызов которой находится на 121 строке.

Тесты:

