1. Из каждого файла 2занятие\_анализ\_данных\_n.xlsx найдите уникальные слова, встречающиеся в файле. Посчитайте количество упоминаний этого слова во всех документах.
2. Методом k-мод выделите 11 кластеров, последний из которых собирал бы “мусор”, имея в виду, что первые 10 кластеров должны иметь осмысленные значения, а 11ый кластер собирать все, что не попало в 11 первых категорий. В качестве расстояния используйте расстояние Левенштейна.
3. Выделите кластер, который содержит наибольшее число элементов.
4. Выделите в указанных документах только заголовки. Проведите действия 1, 2, 3 с заголовками.