***Описание задачи:***

Есть несколько поставщиков услуг (в рамках задачи их 2), которые предлагают в аренду различные номера в различных отелях.

При этом отели по этим поставщикам могут пересекаться.

Мы хотим составить общий доступный номерной фонд для бронирования, но для этого нам нужно провести матчинг между отелями от этих поставщиков.

Поскольку в реальной жизни объемы данных могут быть очень значительными необходимо реализовать алгоритм, который будет выполнять соответствующий матчинг автоматически.

Пример матчинга:

|  |  |
| --- | --- |
| Поставщик\_1 | Поставщик\_2 |
| Atlantis Condo Resort Pattaya By ACC | Atlantis Condo Resort Pattaya by Kulisara |
| Chongqing Yangtze River Internatioanl Youth Hostel | Yangtze River International Youth Hostel |
| Guangdong Hotel Zhuhai | Guang Dong Hotel Zhuhai |
| Hampton Inn Nanuet - NY Hotel | Hampton Inn Nanuet |
| Dong Fang Hotel | Dong Fang Hotel Beijing |

***Цель задания:***

Разработать автоматический алгоритм, который по представленным 2 поставщикам проведет такой матчинг и сформирует соответствующие пары отелей.

Алгоритм должен:

1. максимизировать значение accuracy \* coverage (минимальный порог, которого нужно достигнуть > 0.9)
2. при этом есть ограничение, на accuracy >= 98%

*coverage = correct matches/true matches*

В рамках задания дан только небольшой пример корректно сматченных отелей, истинная разметка не предоставляется.

Качество работы алгоритма будет проводиться по этой истинной разметке.

***Важные комментарии:***

1. можно исходить из того, что большинство отелей (> 99%) первого поставщика имеют соответствующие пары во втором поставщике
2. каждому отелю должен сопоставляться только один другой отель
3. нет требования каждому отелю сопоставить пару
4. матчинг должен проводиться только внутри страны

***Форма предоставления результата:***

Архив, который должен содержать минимум 3 элемента:

1. csv файл с разультатами мапинга в виде:

|  |  |
| --- | --- |
| c1.key | c2.key |
| ACA467BD93827F6001DDD139E7AB4299 | 845F505251CF1D83547D4669A44D17A4 |
| 5BB207CBB33F3D099BCD979F86B3F54B | 9288C09CFB7714C3CAF57723C04BB6A4 |
| 97318B24C9C927AB8F50B13B44A706B5 | 881982B8562A58BF3822C727B8A0C361 |

1. Файл/директорию с исходным кодом алгоритма (читаемый/задокументированный код + how to по запуску, если требуется).

Язык реализации – Python

1. Несколько слайдов в формате power point с описанием подхода, базовых частей решения, проведенным анализов и итоговыми результатами