Отчет о выполнении задания

1. Был создан алгоритм решения задачи:

**Считывание данных** → **Обработка данных** → **Подсчёт метрики** →

→ **Запись данных**

2. Подробнее про каждый пункт:

Считывание данных: Считывание данных из таблиц csv с помощью библиотеки os и csv.

Обработка данных: уменьшение регистра с помощью lower, удаление лишних символов из строки с помощью самостояльно созданой функцией remExc и исправление ошибок и расшифрования аббревиатур с помощью GPT-3.5-TURBO и возможно использование других модель, исправление ошибок лучше делать по штучно во избежания ошибок и аномалий.

Подсчёт метрики: подсчет метрики производится на каждом этапе обработки данных с помощью set – это ускоряет процесс ведь set более быстрый тип данных и не требуется каждый раз использовать массив и проверять каждый его элемент на уникальность.

Запись данных: запись данных производится в таблицу аналогично и для метрики в соответсвии с заданием.

3. Выбор оптимального промпта для GPT:

За время созадания программы были использованны несколько промптов для самого оптимального запроса, путем проб и ошибок был выделен самый лучший:

промпт 1: "Расшифруй сокращения и исправь ошибки."

Выдача GPT:

Директор – глава организации

Помощник руководителя – помощник руководителя организации

Вывод: Данный промпт являлся неэффективным так как выдает не только что требуется, но и лишнию информацию.

промпт 2: "Расшифруй сокращения и исправь ошибки. Примеры: 'ВрИО' → 'временно исполняющий обязанности', 'зам.' → 'заместитель'. Сохрани тире и смысл."

Выдача GPT:

Текст не содержит сокращений или ошибок для расшифровки. Однако, если "Директор" является именем должности, то можно оставить его в текущем виде. Важно уточнить контекст, чтобы внести необходимые изменения.

Помощник руководителя

Водитель автомобиля.

Системный администратор - это специалист, отвечающий за настройку, обслуживание и поддержку компьютерных систем и сетей.

Вывод: Данный промпт являлся более эффективным но не до конца ведь производится выдача лишних символов как точка и подробные записи.

промпт 3: "Расшифруй сокращения и исправь ошибки в

профессиональных должностях. Сохрани тире и смысл. Ответ должен быть в нижнем регистре. Отправь только готовую расшифровку с исправленными ошибками."

Выдача GPT:

Директор

Помощник руководителя

Водитель автомобиля

Вывод: Данный промпт самый эффективный, у данного промпта минимизировались лишние символы и слова.

4. Экономические показатели:

Q – средняя длина строки с должностью

N – количество должностей

C1 – стоимость одного символа в gpt-4.0 на вход

C2 – стоимость одного символа в gpt-3.5-turbo на вход

F1 – стоимость одного символа в gpt-4.0 на выход

F2 – стоимость одного символа в gpt-3.5-turbo на выход

Показатели для gpt-3.5-turbo:

Показатели для gpt-4.0:

Вывод: Цена различается почти в 10 раз, для выопления нашей задачи достаточно будет gpt-3.5-turbo, но при использовании по одному запросу на одну должности иначе выскакивают аномалии которые крайне не желательны при рассшифровке аббревиатур и исправлении ошибок.