У меня есть потребность автоматизировать выявление машинного перевода в тестовых заданиях.

Типа того. Беру оригинал, переводу его через гугл в своём Смарткате, выгружаю перевод. Выгружаю тестовые переводы фрилансеров, они их тоже в Смарткате делают. Поэтому форматирование везде одинаковое. И дальше в ворде в режиме правок сравниваю оригинал и версию фрила. И смотрю "на глаз", сколько там правок и т.п.

Я хочу написать программу для проверки тестовых переводов на предмет использования гугл-транслейт. Менеджеры перевод

В переводческих компаниях всегда большой поток новых переводчиков, которых принимают на работу по результатам тестового перевода – несколько небольших текстов на разные тематики.

В первую очередь нужно отсеять тех переводчиков, которые прислали в качестве тестового перевода результат работы гугл-транслейта. Таких очень много. Я хочу эту работу автоматизировать.

В Когда потенциальный переводчик присылает тестовый перевод, менеджер в первую очередь сравнивает его с результатом гугл-транслейта. Это приходится делать руками – либо в ворде в режиме правок, либо специальными прогами типа copyleaks. Это не очень удобно, поток тестовых переводов большой, каждый тест приходится смотреть отдельно, ворд в режиме правок вообще неудобен и не показателен, copyleaks получше, но он ищет именно плагиат. Переводы одного и того же текста так или иначе похожи, и он между гугл-транслейтом (который едва читаем) и человеческим переводом выдает 40% полных совпадений, например.

В программу можно загрузить тот вариант перевода, который выдал гугл-траслейт, и сразу все файлы, которые прислали переводчики.

гугл с английского на русский художественный текст переводит гораздо хуже, чем с русского на английский что-нибудь технически-юридическое.

Программа выдаёт таблицу, в которой напротив названия каждого файла (в названии фио переводчика) написано, какой процент совпадений. Соответственно, для каждого файла можно открыть и посмотреть, какие там совпадения и какие различия.

Для разных направлений нужны разные результаты.