

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

mlparser (*materials lists parser*) предназначен для извлечения технической информации из таблиц в формате pdf и поиска соответствий в существующей базе данных.

Заказчик, крупное сборочное предприятие Санкт-Петербурга, приобретает у поставщика пакетированные комплекты крепежных изделий — фасованные согласно материальным листам (спецификациям) наборы. Составы комплектов хранятся в таблицах. Ввиду специфичности системы конструкторской документации и политик распространения документов, заказчик предоставляет матлисты только в pdf. Требуется проанализировать файлы в pdf, извлечь из таблиц информацию о составах комплектов и сопоставить номенклатуре заказчика номенклатуру поставщика. Необходимо учитывать, что один артикул заказчика может обслуживаться несколькими наименованиями поставщика, сходными по своим техническим/эксплуатационным характеристикам, но ранжированным по приоритетам коммерческой целесообразности.

Ввиду строгих политик конфиденциальности в отношении материальных ценностей и интеллектуальной собственности вовлеченных компаний, я не могу приложить к данному проекту конкретный пример работы кода, но желающий может самостоятельно прочитать достаточно подробно прокомментированный код.

Основные этапы работы кода:

1. Парсинг матлистов в pdf — извлечение из таблиц данных о составах сборных изделий.
2. Чтение, создание или обновление базы данных соответствия номенклатур заказчика и поставщика.
3. Подбор номенклатуры поставщика для обслуживания новых сборных изделий заказчика с учетом установленных приоритетов.
4. Для удобства менеджера результирующий отчет дополняется условным форматированием: стили для индикации статусов и состояний; валидация предопределенных комментариев и позиций с множественным выбором с помощью выпадающих списков; ссылки на внутренние документы. Предусмотрено распознавание двух наиболее распространенных типов ошибок в документе заказчика с последующими подсказками и предложениями для менеджера поставщика.