support@netology.ru

8 (800) 301-39-69 +7 (495) 152-55-28



Конкурс для Python-разработчиков от компании O2RUS

Задание

Преподаватель: Кристина Байшева

Техническое задание

Создание и заполнение справочника на основе PDF документа

Описание компании:

Транспортная компания «O2RUS» успешно работает на рынке транспортных услуг с 2005 года и имеет репутацию надежного партнера.

Предлагает широкий спектр услуг в организации междугородных и международных перевозок грузов большегрузным автомобильным транспортом на территории РФ и СНГ.

Сайт: https://o2rus.ru/

Описание задачи:

Написать программу на языке Python, которая будет осуществлять парсинг текстового содержимого из PDF документа.

Требования к функциональности:

Программа должна принимать входной PDF файл в качестве аргумента или через интерфейс пользователя, название документа SAE J1939-71.

Программа должна извлекать текстовую информацию из PDF документа и сохранять извлеченный текст в указанный формат базы данных.

Технические требования:

БД - postgresql или sqlite

\equiv

Требования к распознаванию текста:

Программа должна использовать одну из библиотек для PDF-парсинга, таких как PyPDF2, pdfminer.six, PDFMiner или на Ваше усмотрение.

Структура БД:

Таблица с данными

Таблица № 1

Наименование столбика	Описание	Поля с документа	Примечание
ID	Идентификатор пакета CAN шины	ID	
Data_length	Длинна	Data_length	
Length	Длина занимаемая значением	Length	
Name	Наименование	Parameter Name	
RusName	Наименование на русском		
Scaling	Коэффициент	Slot Scaling	
Range	Диапазон	Slot Range	
SPN		SPN	

Описание процесса получения информации с PDF:

 \equiv

- 1. Для заполнения таблицы в начале нас интересует раздел 5.3... документа. Разбегаясь по разделу 5.3... Из данного раздела забираем в таблицу № 1 значения согласно таблице.
- 2. Далее в разделе 5.2.... ищем подробную расшифровку каждого параметра. Поиск производим по 3 признакам. По PGN, названию (Parameter Name в желтой зоне) и Параграфу документа (выделено зеленым). с раздела 5.2 нас интересуют параметры выделенные синим (Рис № 2)

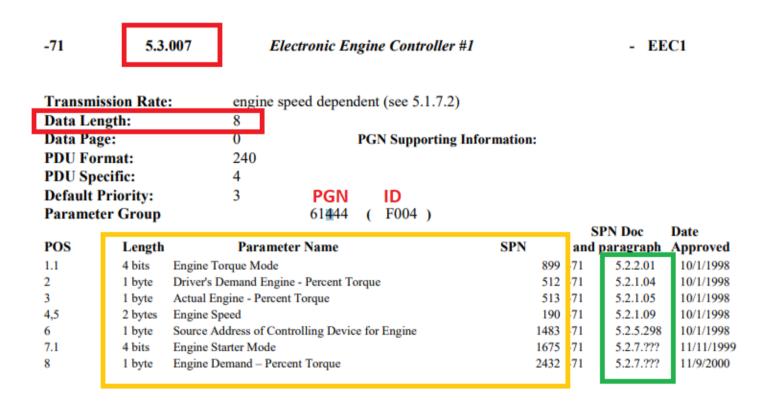


рис № 1 (пример из раздела 5.3)

 \equiv

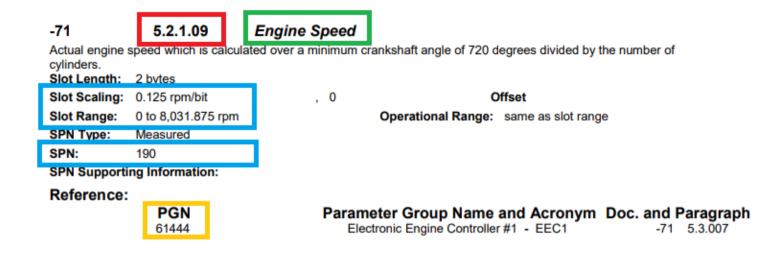


рис № 2 (пример из раздел 5.2)

Требования к документации:

Написать документацию, описывающую функциональность программы, ее использование и возможные настройки.

Ограничения и примечания:

Программа не должна использовать сторонние коммерческие библиотеки или сервисы без предварительного согласования.

Решение:

Ссылка на GitHub с полным описанием REDME и инструкцией по запуску.

Конкурс проводится до 22.02.2024 (включительно)

Отправка задания и вопросы

Отправка задания

Вопросы и ответы

