

Техническое задание

Создание и заполнение справочника на основе PDF документа

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.

[Образец pdf файла.](#)

Технические требования:

БД - postgresql или sqlite

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:

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)

-71	5.3.007	Electronic Engine Controller #1	- EEC1
Transmission Rate:	engine speed dependent (see 5.1.7.2)		
Data Length:	8		
Data Page:	0	PGN Supporting Information:	
PDU Format:	240		
PDU Specific:	4		
Default Priority:	3	PGN ID	
Parameter Group	61444	(F004)	
POS	Length	Parameter Name	SPN
1.1	4 bits	Engine Torque Mode	899
2	1 byte	Driver's Demand Engine - Percent Torque	512
3	1 byte	Actual Engine - Percent Torque	513
4.5	2 bytes	Engine Speed	190
6	1 byte	Source Address of Controlling Device for Engine	1483
7.1	4 bits	Engine Starter Mode	1675
8	1 byte	Engine Demand – Percent Torque	2432

рис № 1 (пример из раздела 5.3)

-71	5.2.1.09	Engine Speed	
Actual engine speed which is calculated over a minimum crankshaft angle of 720 degrees divided by the number of cylinders.			
Slot Length:	2 bytes		
Slot Scaling:	0.125 rpm/bit		Offset
Slot Range:	0 to 8,031.875 rpm		Operational Range: same as slot range
SPN Type:	Measured		
SPN:	190		
SPN Supporting Information:			
Reference:	PGN 61444	Parameter Group Name and Acronym Electronic Engine Controller #1 - EEC1	Doc. and Paragraph -71 5.3.007

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 README и инструкцией по запуску.