ANALYSIS OF THE DATA EXTRACTED FROM AN ENGINEERING DRAWING

A PROJECT REPORT

Submitted by

PARTHASARATHI ROY

in partial fulfilment for the project

CAD ANALYTICS (CO26)

under the

DATA SCIENCE INTERNSHIP

IDEAS - INSTITUTE OF DATA ENGINEERING, ANALYTICS AND SCIENCE FOUNDATION

TECHNOLOGY INNOVATION HUB @ INDIAN STATISTICAL INSTITUTE, KOLKATA

NOVEMBER 2024

TABLE OF CONTENTS

TITLE	PAGE NO.
INTRODUCTION	1
METHODOLOGY	2
PROBLEM STATEMENT	3
WORK APPROACH	4
1. AutoCAD data extraction	4
2. Analysis of the extracted data	6
3. Emphasis on the technical aspects	8
CONCLUSION	10
WORK PRODUCTS AND DELIVERABLES	11

INTRODUCTION:

CAD Analytics refers to the process of extracting, analyzing, and interpreting data from Computer-Aided Design (CAD) files (like DWG, DXF). It leverages analytical tools and algorithms to gain insights from CAD drawings and models for better decision-making in engineering, manufacturing, architecture, and other related fields. CAD Analytics bridges the gap between design and data analysis, offering powerful insights to improve efficiency, cost, and accuracy in projects across various industries.

METHODOLOGY:

In this project we have predominantly worked with 2D CAD files (DWG, DXF and PDF files) i.e. engineering drawings and tried to follow the methodology given below:

- **Data Extraction**: Extract geometric, annotation, and metadata information from CAD files.
- **Metadata Analysis**: Analyse layers, dimensions, units, materials, and object properties.
- **Dimensional Analysis**: Measure lengths, areas, volumes, and perform geometric calculations.
- **Text Extraction**: Extract part lists, quantities, and materials from CAD drawings.

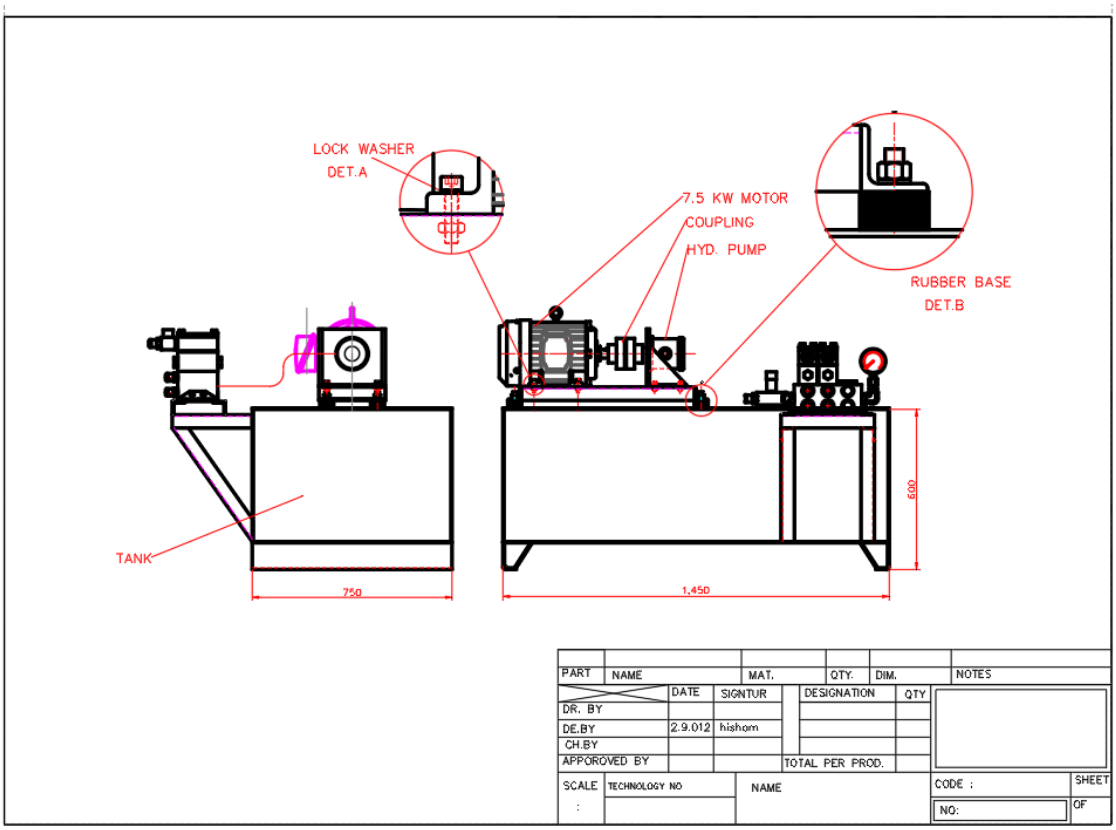
PROBLEM STATEMENT:

- Massive amount of data present in the engineering drawings.
- Important technical information about the drawing like dimensions, assembly part data and operational parameters are missing.
- Data extracted from the drawings using online APIs is generally in XML or JSON format which is not clearly understandable.

WORK APPROACH:

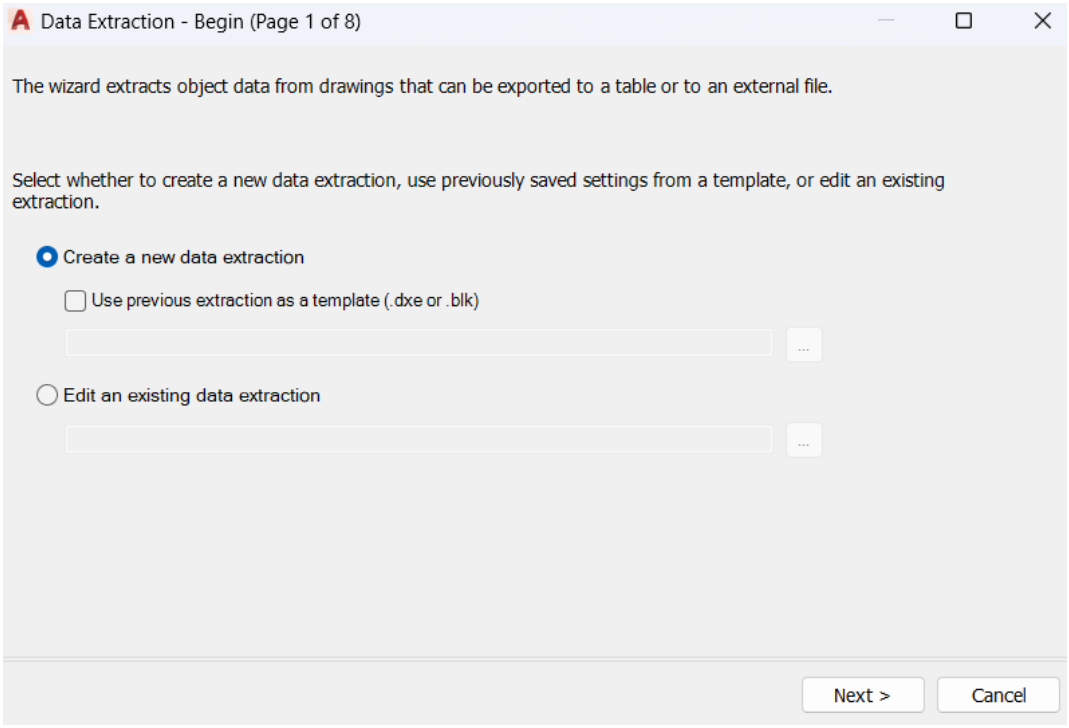
1. AutoCAD Data Extraction

A file named hydraulic_unit.dwg is opened in AutoCAD.

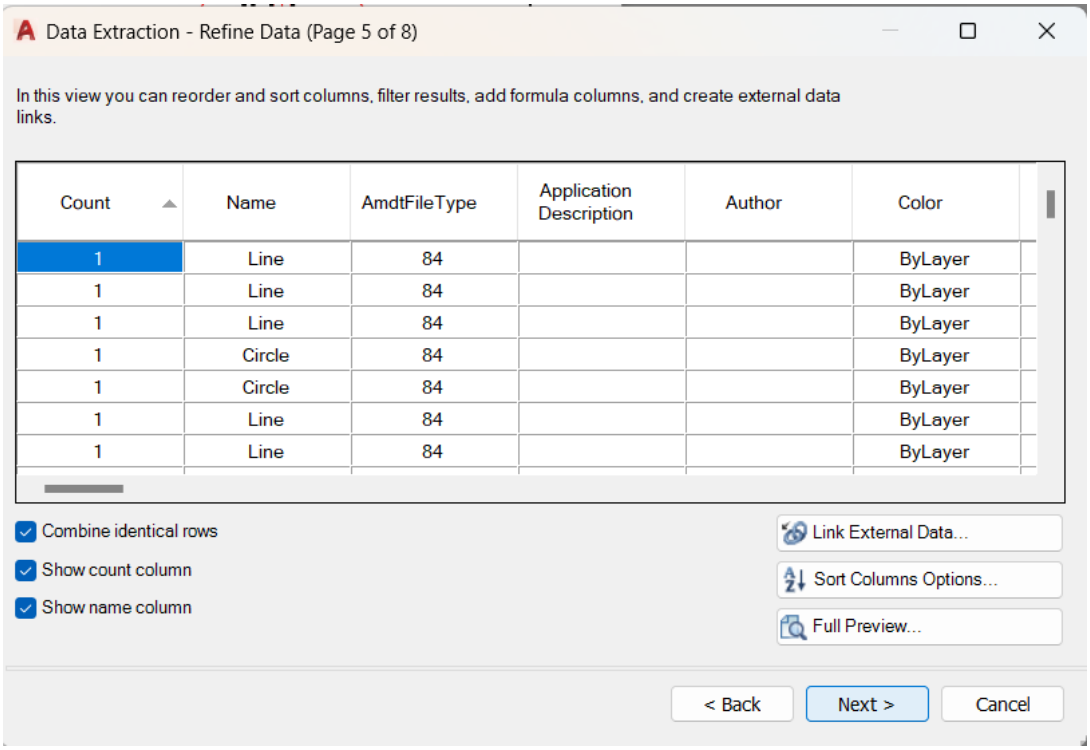


On the insert tab, there is a panel called “Linking and Extraction”. One of the tools available on this panel is Extract data. Once this tool is selected the data extraction dialog box will appear.

The first step is to create a new data extraction and eventually follow the next steps in the data extraction process.



Once we reach the Refine Data view, we can reorder and sort columns and filter results as per our needs.



At last export the data as .xls file and eliminate unnecessary columns to show only the required attributes.

2. Analysis of the extracted data

After manually comparing the extracted data with the drawing, it has been observed that there has been some unnecessary data. After removing some empty columns, the data still looks complicated.

Count	Name	Color	Layer	Linetype	Linetype Scale	Lineweight
1	Line	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Line	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Line	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Circle	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Circle	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Line	ByLayer	MOTOR	ByLayer	1.0000	ByLayer
1	Line	ByLayer	AM_0N	ByLayer	1.0000	ByLayer
1	Line	ByLayer	AM_0N	ByLayer	1.0000	ByLayer
1	Arc	ByLayer	AM_0N	ByLayer	1.0000	ByLayer
1	Line	red	0	HIDDEN	0.2500	0.20 mm
1	Line	red	0	HIDDEN	0.2500	0.20 mm
1	Line	ByLayer	AM_0N	ByLayer	1.0000	ByLayer
1	Arc	magenta	MOTOR	ByLayer	1.0000	0.40 mm
1	Line	magenta	MOTOR	ByLayer	1.0000	0.40 mm
1	Line	magenta	MOTOR	ByLayer	1.0000	0.40 mm
1	Line	magenta	MOTOR	ByLayer	1.0000	0.40 mm

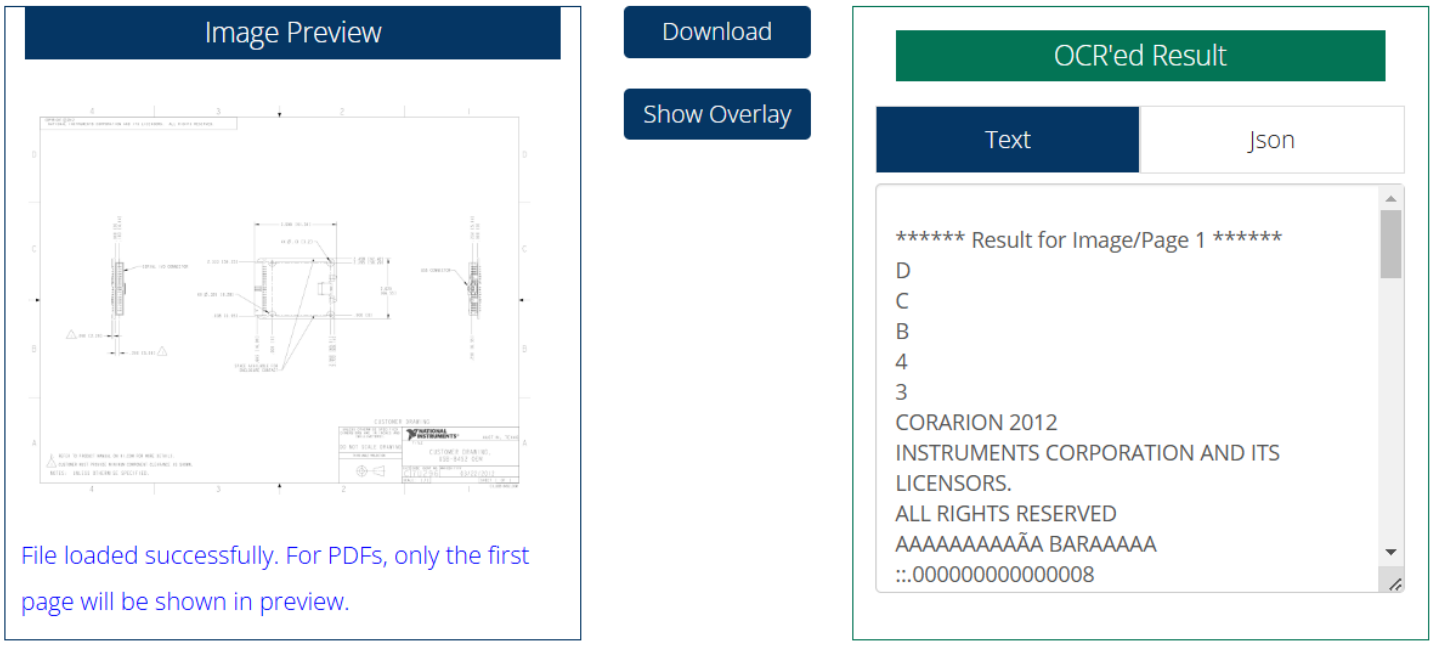
Plot Style	Area	Center X	Center Y	Length	Radius	Start Angle	Total Angle	Circumference	Diameter	Angle
ByLayer				0.2854						270
ByLayer				0.0000						0
ByLayer				26.7800						90
ByLayer	2253.0510	-1607.0738	-2293.1426		26.7800			168.2637	53.5600	
ByLayer	2355.1402	-1607.0738	-2293.1426		27.3800			172.0336	54.7600	
ByLayer				0.8556						90
ByLayer				9.0000						270
ByLayer				14.0000						0
ByLayer	0.2854	-1709.3561	-2398.2402	1.5708	1.0000	90	90			
Color_1				16.0000						270
Color_1				16.0000						270
ByLayer				9.0000						270
Color_6	0.1202	-1607.0738	-2293.1426	5.2806	102.0600	99	3			
Color_6				2.9890						264
Color_6				2.9890						296
Color_6				2.9855						100
Color_6	1.4217	-1607.0738	-2293.1426	11.9272	99.3800	97	7			
Color_6	0.0383	-1607.0738	-2293.1426	3.6402	104.9160	99	2			
Color_6				0.7094						270
Color_6				2.9820						100
Color_6				2.9820						100
Color_6	2.7581	-1627.3077	-2194.4717	4.1635	1.3491	277	177			
Color_6	2.7581	-1621.8076	-2193.5019	4.1635	1.3491	106	177			
Color_6	8.2140	-1607.0738	-2293.1426	21.5144	100.8000	101	12			

Delta X	Delta Y	End X	End Y	Start X	Start Y
0.0000	-0.2854	-1607.0738	-2319.9226	-1607.0738	-2319.6371
0.0000	0.0000	-1607.0738	-2319.9226	-1607.0738	-2319.9226
0.0000	26.7800	-1607.0738	-2293.1426	-1607.0738	-2319.9226
0.0000	0.8556	-1607.0738	-2319.6371	-1607.0738	-2320.4928
0.0000	-9.0000	-1710.3561	-2407.2402	-1710.3561	-2398.2402
14.0000	0.0000	-1695.3561	-2397.2402	-1709.3561	-2397.2402
0.0000	-16.0000	-1707.6561	-2425.4402	-1707.6561	-2409.4402
0.0000	-16.0000	-1697.0561	-2425.4402	-1697.0561	-2409.4402
0.0000	-9.0000	-1694.3561	-2407.2402	-1694.3561	-2398.2402
-0.3084	-2.9731	-1627.3903	-2193.1252	-1627.0819	-2190.1521
1.3067	-2.6883	-1622.1905	-2192.2083	-1623.4972	-2189.5200
-0.5184	2.9402	-1630.1694	-2196.4835	-1629.6509	-2199.4237

Contents	Position X1	Position Y1	Width	Position X	Position Y	Closed	Height	Style	Value
				408.9347	55.7087		2.5000	STANDARD	QTY
				328.5614	44.4347		2.5000	STANDARD	2.9.012
				345.5700	44.4347		2.5000	STANDARD	hisham
TOTAL PER PROD.	367.4021	34.7136	37.8299						
				419.2344	25.0948		2.5000	STANDARD	CODE :
				355.4602	62.4131		2.5000	STANDARD	MAT.
				383.4428	62.4131		2.5000	STANDARD	QTY.
				467.3906	26.3953		2.5000	STANDARD	SHEET
				466.7398	17.9424		2.5000	STANDARD	OF
						-1			
						-1			
				421.1867	15.9917		2.5000	STANDARD	NO:

The attributes that we can see above are Count, Name, Color, Layer, Linetype, Linetype Scale, Lineweight, Plot Style, etc.

An alternative approach can be taken by converting the drawing file (DWG/DXF) to PDF files or downloading the drawings directly in PDF format. Websites like https://ocr.space/ has been used to extract the text data from this PDF which might be easily readable and user friendly.

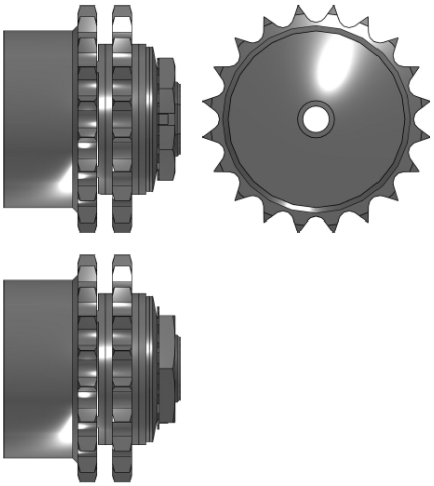


3. Emphasis on the technical aspects

We referred to OEM catalogues in https://www.3dfindit.com/en/ for better understanding of the technical data in the OEM drawings

Let us take an example of a Slip Clutch Hub

2D derivation

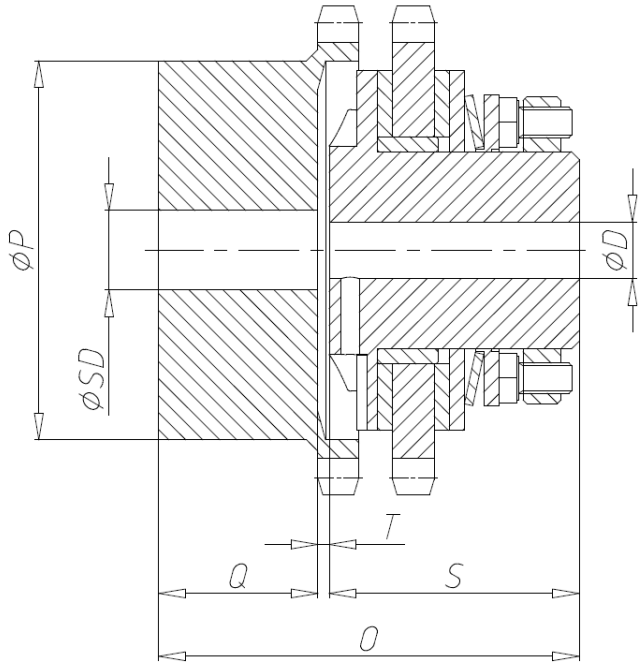


ISO view

Bill-Of-Materials

N°	Description	Amount
1	Hatorq-slip hub couplings KETTEN FUCHS Artikel-Nr. C 240/85	1

Technical drawings



Technical data

CNSORDERNO (Order number)	C 240/85
MTA (Size Mt × A)	240 × 85
MT (Transmission values max torque Mt. / Nm)	240
AT (Number plate spring)	2
DP (Parallel offset Δp)	0.35
AW (Axial offset αw / Grad)	30'
SD (Bore sprocket hub / mm)	21.0
SDB (sd min. / mm)	21.0
SDMAX (sd max. / mm)	65.0
D (Wheel hub bore chains / mm)	15.0
DB (D min / mm)	15.0
DMAX (D max / mm)	25.0
O (Dimensions / mm)	100.0
P (Dimensions / mm)	100.0
Q (Dimensions / mm)	42.0
S (Dimensions / mm)	55.0
T (Dimensions / mm)	3.0
Z (Number of Teeth)	20
KG (Chain size)	3/4 × 7/16"

The above table clearly shows the technical data about the OEM in an organized manner. Hence a similar approach has been undertaken for the AutoCAD data extraction of the above drawing file and the results are below.

	A	B	C	D	E	F	G	H
1	AT	AW	CATALOG	CNSORDERNO	Comments	COMPANY	D	DB
2	2	30?	Ketten Fuchs	C 240/85	File was created by Cadenas	ketten_fuchs	15.0	15.0
3								

	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	DMAX	DP	KG	Layer	MT	MTA	O	P	Q	S	SD	SDB	SDMAX	T	Z
2	25.0	0.35	3/4 × 7/16"	AM_6	240	240 × 85	100.0	100.0	42.0	55.0	21.0	21.0	65.0	3.0	20
3															

CONCLUSION:

There are numerous APIs already available which can extract data from engineering drawings generally in XML or JSON format. There are also native CAD tools like AutoCAD which can also extract the data from engineering drawings in Excel format using its Data Extraction Wizard. The data which is extracted in both these scenarios seems to be complicated from the perspective of a customer.

The possible solution is to develop something that can breakdown these massive data into smaller chunks and organize it in order to increase the efficiency in the early stages of the Product Lifecycle Management (PLM).

WORK PRODUCTS AND DELIVERABLES:

Software used for Data Extraction: AutoCAD 2019

Links to Original Equipment Manufacturers' (OEMs) CAD Files:

1. Siemens
2. Eaton
3. ABB
4. HermanMiller
5. Kohler
6. Mitsubishi Electric
7. Schneider Electric
8. Saint Gobain

Websites of CAD (Computer Aided Design) Communities:

1. CAD Forum
2. 3Dfindit
3. 3dcontentcentral

APIs used for Metadata Extraction:

1. Aspose
2. GroupDocs
3. OCR.space

Github Repository Link:

CAD-Analytics