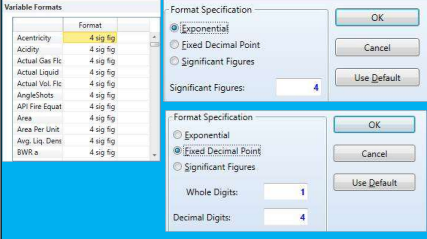


PDF Data전산화

수집					분석		평가			
NO	*대분류	*중분류	*요구사항상세	*세부 항목	평가 항목 분석	구현분석	*우선순위 (IT관점)	*범위승인 (Y/N)	비고 (고려사항, 제외사유 등)	
1.1	전처리 및 Migration Data	Migration	대림의 기수행 Project Datasheet 의 데이터는 모두 탑재되어야 한다. 개발시 대림직원이 업로드 DB Mapping 테스트 및 개발시 - 기수행 Project 는 42 (?) 정도 의 Project 를 대상으로 함.	-	2.2.3으로 변경	N/A	-	N	개발 요구사항 아님(DL Scope)	
2.1.1	Solution 기능	추출 기술	Native File 형태(Converted PDF, Excel, Scanned PDF, Autocad, XML, EXCEL) - 개별 or Multi 파일 Upload / Drag & Drop 방식	Converted PDF	각 세부 항목별 우선순위 여부, 개발 필수 항목 유/무 등 확인 필요					
				Excel						
				Scanned PDF						
				Autocad						
				XML	상용 SW 의 Output XML 파일로써, DB 속성과 Mapping 고정					
EXCEL	In-house Excel Program 파일로써, DB 속성과 Mapping 고정									
2.1.2			Data 추출 영역 기준(Table, Note, Sketch) - 하나의 Cell에서 분리 추출(특수문자등으로 구분 시에도 이를 분류하여 추출 ex. Design Pressure: 38.72 kg/cm2 (See Note. 1) 인 경우, 38.72 (Value), kg/cm2 (Unit), (See Note. 1) 이 각각 구 분되어 추출되어 Design Pressure 라는 속성명이 DL 표준 속성 과 Manninng 될 수 있어야 함 - Remark / Note 등으로 구분된 영역의 Sentence 가 문장 혹은 Line 단위로 추출 - Image 추출 기능 (Image내 Data추출 기능) - 추출방식: Dictionary 와 Matching, Area 설정 등 1. 추출대상: Datasheet의 Nozzle Table 영역 / Tray Loading Data / Heat Curve 등의 영역 2. 형식: Value (숫자) + Text (문자) 3. 추출-Mapping 기능과 방법 - 아래 참조 인식되어야 하는 개수를 파악하고, 해당 표준속성과 Data가 1대1 매 칭이 아닌 1대다 매칭으로 인식되어야 함. ex) Nozzle이 10개인 Datasheet 표준속성명 : Nozzle No. (1개) 매칭 Value : N1..N2..N3...N10 (10개)	Table	각 세부 항목별 우선순위 여부, 개발 필수 항목 유/무 등 확인 필요					
				Note						
				Note (Sentence)						
				Sketch						
					개발사 제시					
2.1.3			속성의 형태(Value, Text, Unit(UOM), 기호) - Unit Defintion에 저장된 항목중 동일한 항목 추출함 - 대문자, 소문자, 위,아래 첨자등을 구분 가능함 - 기호(●,○,■,□)를 Yes, No등으로 Data화 가능	Value Unit(UOM) Text 기호	확인 사항 : 소수점 구분					
2.1.4			Document 유형(Datasheet, HMB, VP, Drawing(G.A., P&ID, ISO)) 개발우선순위는 1순위 Datasheet, 2순위 기타문서로 추진함	Datasheet	각 세부 항목별 우선순위 여부, 개발 필수 항목 유/무 등 확인 필요					
				VP Drawing						
2.2.1		Setup	표준 속성명 Setup - DL 표준 속성명 정의(입력/조회/수정)							

REACTOR DATA SHEET						
NOZZLE CONNECTIONS (I)						
SYM	SIZE	QTY	RATING	FACING		SERVICE
N1A	1 1/2X12	1	2000	LWN RTWLN		CATALYST INLET
N2B	1 1/2"	1	2000	LWN LN		CATALYST FEED (W/INSTR) TO W/INSTR
N2	1 1/2X12	1	2000	LWN RTWLN		TOTAL FEED TO CSTR
N3	8"	1	2000	WLN RP		LOOP INLET
N4	8"	1	2000	WLN RP		LOOP CHILET
N5	4"	1	2000	PAD L F		EMERGENCY DUMP
N6	1/2"	1	2000	PAD L F		INSTRUMENT
N7	2"	1	2000	PAD RP		PRESSURE SAFETY VALVE
N8	1/2"	1	2000	PAD RP		INSTRUMENT

2.2.2		속성 Dictionary(Library) Setup - 표준 속성명과 대응하는 유사, 유의어 등을 정의(입력/조회/수정)						
2.2.3		속성 Dictionary(Library) Data구축 - 당사 기 수행 Project Document의 속성명을 추출하여 Dictionary구축 : (개발사) 전달된 Document 에서 속성명 Data를 추출하여 전달 : (DL E&C) 추출된 속성명들 DL 표준 속성명과 Matching 작업 (개발사에서 이를 쉽게 수행할 수 있는 방안을 제시) - DL 표준속성의 선택형 Data 의 경우, 추출된 속성에서 선택형 Data 와 사전 Matching 작업	(참고사항) DL 표준 속성은 크게 Common 분류와 Equipment Type 별 분류로 나뉜. - Common 분류: 모든 EQT 에 공통 적용 - Equipment Type 별 분류: 특정 EQT 에서만 생성되는 속성 - 각 분류 하위로 Process Condition / Duty 와 같은 Group 이 이루어지며, 해당 Group 별 Data 속성명이 정의되어 있음. - 분류 Level 은 현재 3 level 로 구성 예정이나, 4 level 로 in-depth 한 정의가 필요할 수 있음					
2.2.4		Unit Definition Setup - Master DB에 생성될 모든 UOM을 별도 저장 관리(입력/조회/수정) - Unit 별 표기할 소수점 자리수 Setting - Unit 을 활용하는 Data 속성을 DL 표준속성에 따라 개별 선정 및 DL 표준 속성명을 Grouping 하여 적용 - 선택형 Setup - 추출된 Text / 기호 항목을 DL 표준속성의 선택형 Text 값과 Matching 할 수 있는 Setuo 기능 제공 - 타 프로그램과 같이 유효숫자 (Significant Figures) 개념을 도입하여, 최소한 아래 3가지 형태를 참조하여 반영할 수 있도록 구성.						
2.2.5		Template 정의 Setup - 파일의 유형 별 Template정의						
2.2.6		Dictionary의 ISO 15926 Dictionary 제공						
2.3.1	Mapping	자동 Mapping 기능 - 페이지 인식 후 정의된 Dictionary를 기준으로 자동 Mapping EQ Datasheet (Converted PDF 기준) 우선으로 개발 - Mapping 목표 성공률 (인식 & Mapping 정확도) - Converted PDF 기준 2021년 70% (ITEM 별로 나누어 측정) 2022년 상반기 80%, 하반기 90% 최근 Proposal이용 Datasheet Propety 수량중, 인식 & Mapping 성공율 기준으론 향상						
2.3.2		자동 Mapping 기능 - Value 는 Unit 과 연계되어 Mapping 되어야 함 (Unit 이 상이한 경우, Value 값이 Conversion 되도록) - 추출된 Text 가 DL 표준 속성의 선택형 Value 인 경우, 사전 Setting 된 결과에 따라 DL 표준 속성의 선택형으로 전환되어 선택되어야 함						
2.3.3		수동 Mapping 기능 - 자동 Mapping완료 후 잔여 분에 대한 수동 Mapping기능						
2.3.4		Mapping 편의 기능 - 프로그램에서 구획을 자동으로 나누고 (데이터 집합단위), 구획구분을 유저에게 추천함 - 하나의 페이지를 구획별로 나눠서 Mapping하는 기능						
2.3.5		Mapping 편의 기능 - Mapping완료 후 결과 조회 화면(Preview)						
2.3.6		Unit Conversion 기능 - 추출된 Value 의 Unit 과 Master DB Unit 이 상이한 경우, 이를 자동으로 Conversion 할 수 있는 기능 자동 Mapping 이 되지 않은 Dictionary 의 경우, 사용자의 수동 Mapping 경향을 학습함으로써 사용할수록 수동 Mapping 이 자동 Mapping 으로 전환될 수 있는 Machine Learning 기능 탑재 (Option) 또는 향후 ML 개발이 별도로 가능하도록 학습용 Data 축적	시관련 사항은 후속과제로 고려 필요	-	N		향후 개발	

