

AI-driven IT Automation for Enterprise Business

August 24, 2022

권영준

SAMSUNG SDS

AI for Business

- 오늘의 business application & IT process에는 단순하거나 지능을 요구하는 반복적 수작업(human labor)의 개입이 불가피
- AI-driven IT Automation \triangleq AI 기술을 통해 비효율적 요소와 수작업을 최대한 줄이는 자동화

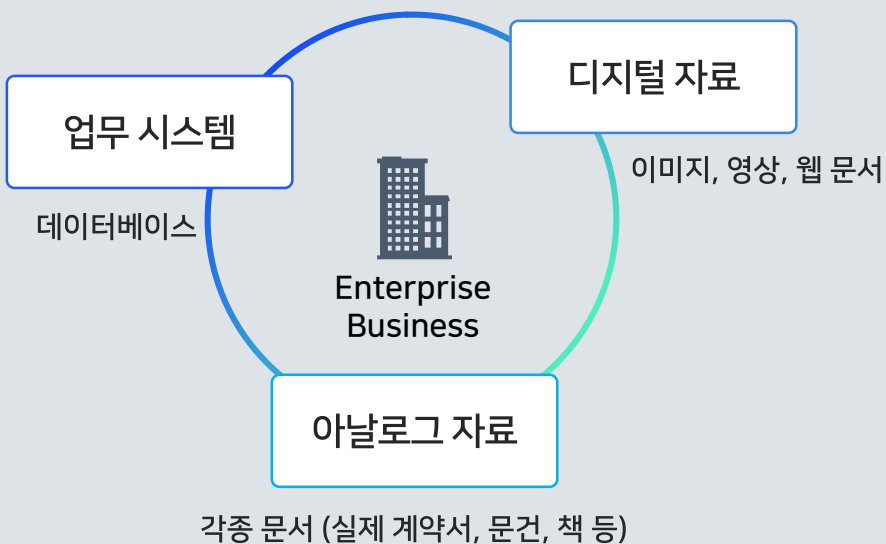


기업 경쟁력 강화를 위해 IT automation 을 통한 비용절감, 생산성 향상 니즈가 증가함에 따라
이를 지원하는 AI 기술이 중요해짐

중요 AI 기술

IT automation을 위해서는 enterprise 내 존재하는 웹, DB, 이미지, 동영상, 디지털-아날로그 문서 등 각종 데이터에 존재하는 정보를 통합 분석하고, 언어를 이해하는 Linguistic Intelligence와 이미지를 이해하는 Visual Intelligence가 중요

Enterprise Business Data



AI

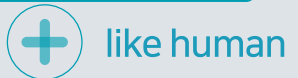
Linguistic Intelligence

- 자연어 이해를 바탕으로 기업 내 정형, 비정형, 반정형의 다양한 text 정보를 사람의 사고 방식처럼 이해하고 분석

Visual Intelligence

- 이미지나 인쇄 문서에서 글자를 이해하고 digital화
- 각종 제조결함을 분석하는 등 다양한 시각적 처리를 진행

멀티모달 데이터 통합



정보와 프로세스 이해



프로세스 자동화



비즈니스 혁신

삼성SDS AI Technologies for Business

Applications

VoC, MRC, Semantic Search, AICR, DCD, Combinatorial Optimization

핵심기술

Linguistic Intelligence

Natural Language Modeling,
기계독해 (MRC, QA),
Knowledge Graph, Sentence &
Paragraph Embedding,
Summarization, ...

Visual Intelligence

Optical Character Recognition,
Document Change Detection,
Object Classification, Face
Recognition, 3D Vision, Defect
Analysis, Explainable AI, ...

선행기술

Reinforcement Learning, Meta Learning, Quantum Computing

Natural Language Understanding

AI 딥러닝 모델의 한계는 scale up이며 이를 위해 스스로 자동으로 문제를 만들고 학습하는 self-supervised learning (자기지도 학습) 방법으로 pretrain, 추후 실제 문제에 fine-tune하여 다양한 도메인의 태스크를 수행

Open-source BERT 기반으로 in-house 한글 모델 개발

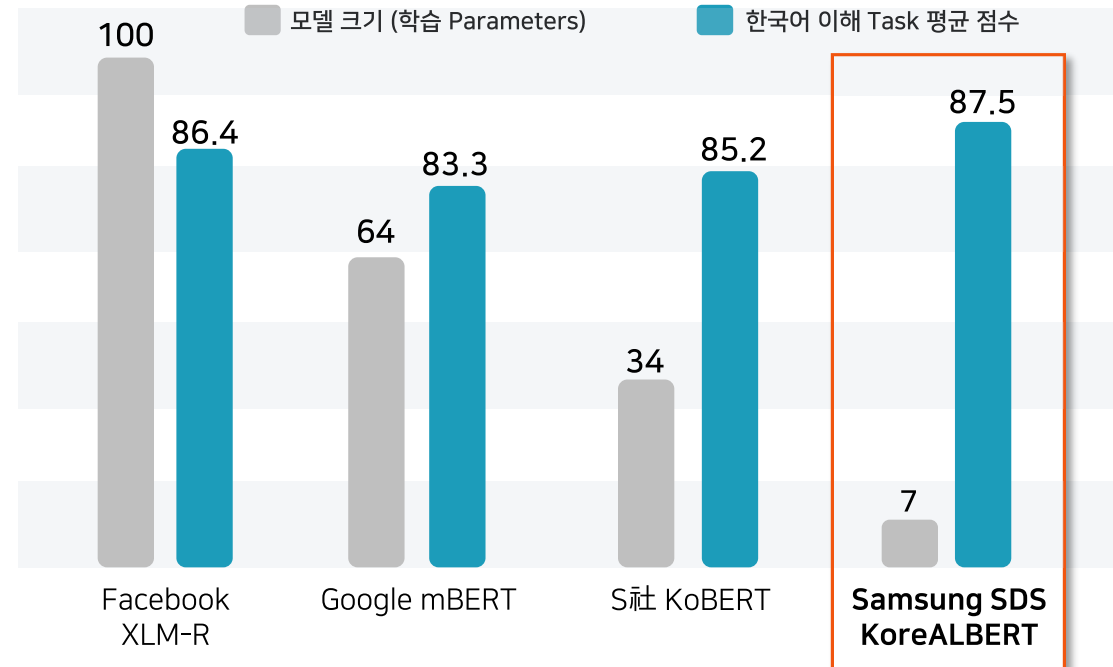
SDS KoreALBERT

- 정확도 손실이 거의 없는 경량 모델 (BERT 대비 1/10 크기)
- 학습에 적은 수의 GPU 필요
- CPU 에서 추론 가능

SDS KoELECTRA

- KoreALBERT 보다 크지만 더 좋은 성능
- 다양한 용도와 domain에서 적용 가능
- 한영문 혼용문장에도 높은 정확성 유지

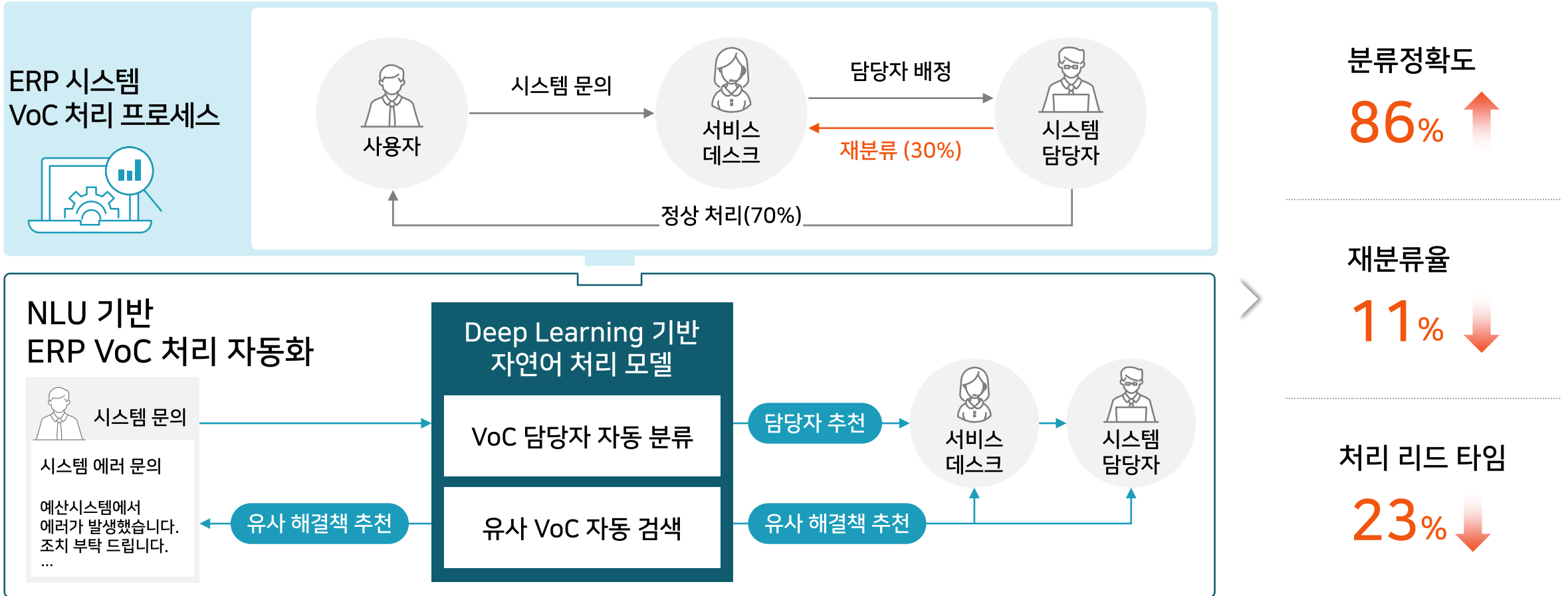
주요 언어모델 성능 비교



한글 기계독해 (MRC) 리더보드 KorQuAD 1.0 & KorQuAD 2.0 동시 석권

Case Study 1: VoC

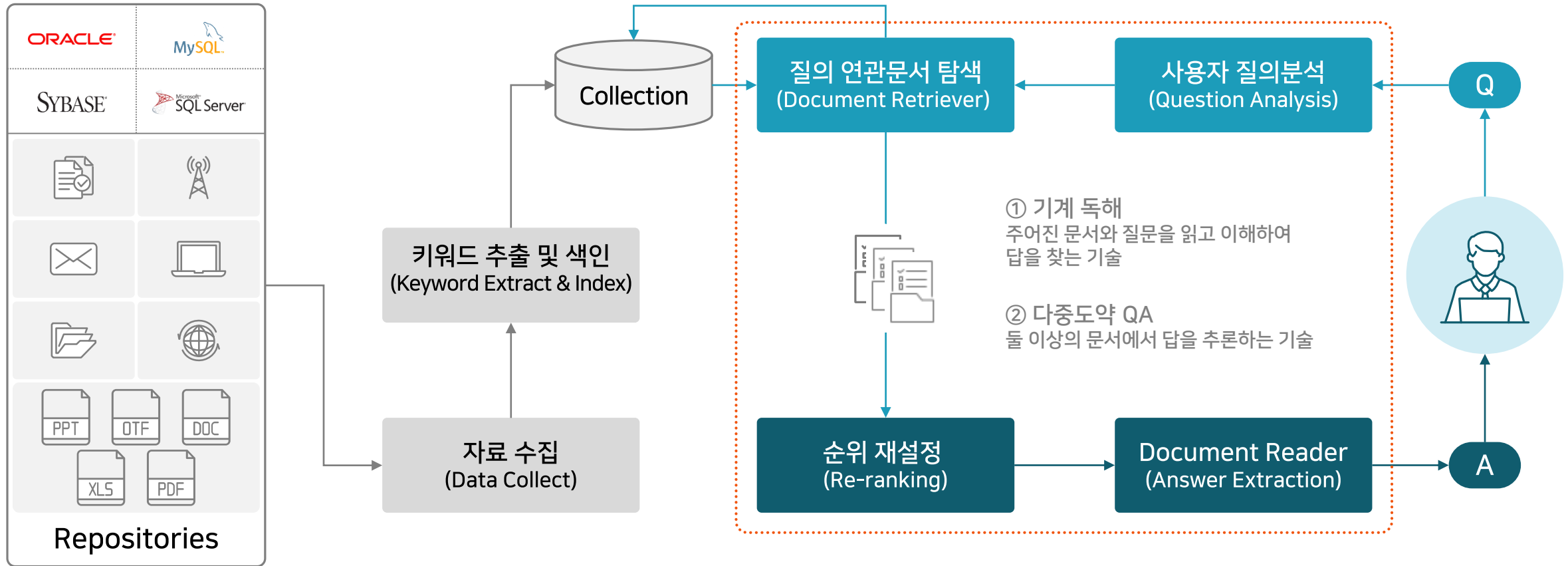
ERP 시스템의 VoC (Voice of Customer) 처리 프로세스에서 수작업 비효율성을 개선하기 위해 pretrained language model을 활용한 자연어 처리로 재분류율을 낮추고 전체 처리 리드타임 단축



VoC (Voice of Customer): 시스템 단순문의, 긴급장애처리, 시스템 기능 변경 요청, 시스템 신규 기능 개선 요청

Case Study 2: 기계독해 QA

일반검색(lexical search)을 넘어서 문장이나 문단을 이해하여 질의응답 (Question-Answer, QA)이 가능한 AI 기계독해 (MRC) 기반의 semantic search를 Enterprise Search Engine의 기능으로 탑재



Case Study 2: 기계독해 QA (cont.)

순수 텍스트 외에 문서 내 포함된 테이블이나 리스트 형태의 정보도 기계독해 QA 가능

테이블, 리스트 구조 MRC & QA

Question: 2016시즌 부천FC는 4강에서 누구에게 패배했나?

Answer: FC서울

Support: 한글 Wikipedia 부천FC 관련정보(테이블 포함)

라운드	날짜	상대팀	경기장	스코어	경기결과
3라운드	4월 27일	한라대학교	부천종합운동장	3:1	승
32강	5월 11일	포항 스틸러스	포항스틸야드	2:0	승
16강	6월 22일	경주시민축구단	부천종합운동장	3:1	승
8강	7월 13일	전북 현대 모터스	전주월드컵경기장	3:2	승
4강	10월 26일	FC 서울	서울월드컵경기장	0:1	패

삼성SDS 기계독해 QA 기술

- ☑ 딥러닝 기반 MRC 기술 개발
→ 한국어 MRC competition인 KorQuAD 2.0 1위 ('20.5)
- ☑ 딥러닝 기반 Multi-hop QA 기술 개발
→ Global competition인 HotpotQA 1위 ('20.9)
- ☑ 딥러닝 모델 기반 검색 고도화 기술 개발
→ Global competition인 Microsoft MARCO 1위 ('21.3)

Case Study 3: 지식그래프

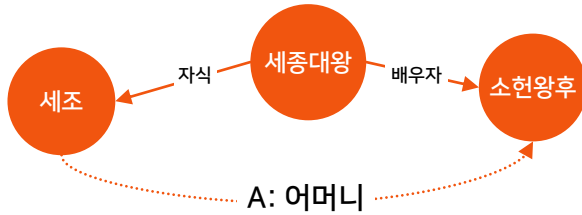
지식 그래프는 기존 데이터베이스 기술에 비해 직접 명시되지 않은 관계에 대한 추론이 가능하고, 복잡한 참조(multi-hop)가 필요한 QA에 장점, 데이터 추가/삭제/변경 시에도 관리가 용이

지식그래프

기존서비스

Reasoning

Q: 세조와 소헌왕후의 관계는?



관계 추론 가능

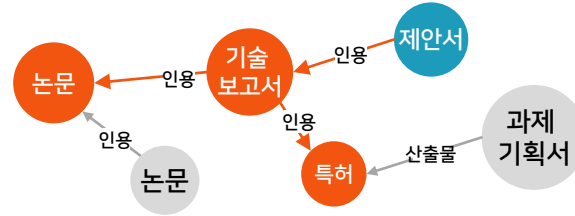
관계 테이블

인물 A	인물 B	관계
세종대왕	소헌왕후	배우자
세종대왕	세조	자식

정의되지 않은 관계 유추 불가능

Performance

Q: 제안서가 인용한 문서가 인용한 문서는?



3회 그래프 액세스

문서 테이블

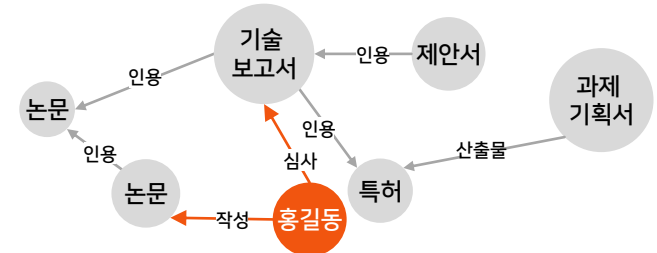
ID	Doc	DocID	RefID
1	제안서	1	2
2	기술 보고서	2	3
3	특허	2	5
4	논문	4	6
5	논문	5	4

2개 테이블 12회 데이터 액세스

인용 테이블

DocID	RefID
1	2
2	3
2	5
4	6
5	4

Flexibility



간편한 데이터 추가/삭제/변경

문서 테이블

ID	Doc
1	제안서
2	기술 보고서
3	특허
5	논문

인용 테이블

DocID	RefID
1	2
2	3
2	5
5	6

심사 (새 테이블)

PER ID	RefID
6	3

인물 (새 테이블)

ID	이름
6	홍길동

작성 (새 테이블)

PER ID	RefID
6	5

데이터 변경에 복잡한 과정 필요

Case Study 3: 지식그래프 (cont.)

SDS 지식포탈 아리샘에 지식 그래프 기반 검색 고도화 적용

ARISAM Nexledger

검색필터 초기화

지식카테고리

- 마케팅(18)
 - Br
 - Ar
- 상품관리(206)
 - Sa
 - 상
 - Ri
- 개발(44)
 - St
 - Te
 - 재
 - 오
 - U
 - 개

등록자 등록일 제목

통합검색 아리샘의 통합 검색은 회사내 지식사이트를 통합하여 검색합니다. 운영필자

Nexledger에 대한 검색결과 (307 건)

정확도 | 최신순 ☐ 제목만 보기

Nexledger
마케팅
등록자
- ai : 전시 행사 등 그래픽 디자인서 사용 - pdf : 영어용 - ppt : 보고자료, 발표자료 등 문서작성시 사용 ※ 기타문의 : 글로벌마케팅그룹 운영중 프로 ※ 이미지 파일은 제공한 원본 그대로 사용하고 별도 추가/수정 등 변경 금지. Please use the logo as provided, do not add or modify the...

[PDF] Nexledger
마케팅
등록자

[AI] Nexledger
마케팅
등록자

[Brochure] Nexledger
상품관
등록자
[Brochure] Nexledger Contact Information -

[PPTX] Nexledger
마케팅
등록자

Nexledger 에서 BVA란?
개발 >
등록자
Nexledger 에서 BVA(BlockchainValue Assessment)라는 적용 방법론이 있는 것으로 알고 있습니다. 자세한 내용을 알고 싶는데 어디로 문의하면 되나요?

Nexledger v2.2 Brochure
상품관리
등록자 :
Nexledger v2.2 Brochure(KR/EN)

[PPTX] Nexledger Universal

Nexledger 개선요청

키워드 정의

- Nexledger : SDS가 보유한 Blockchain 솔루션
- * Blockchain : 지속적으로 변경되는 데이터의 변경이력을 모든 참여주체가 저장하여, 누구도 임의로 수정할수 없으며 누구나 조회할수 있는 네트워크이자 데이터베이스

연관 키워드 기반 탐색 ①

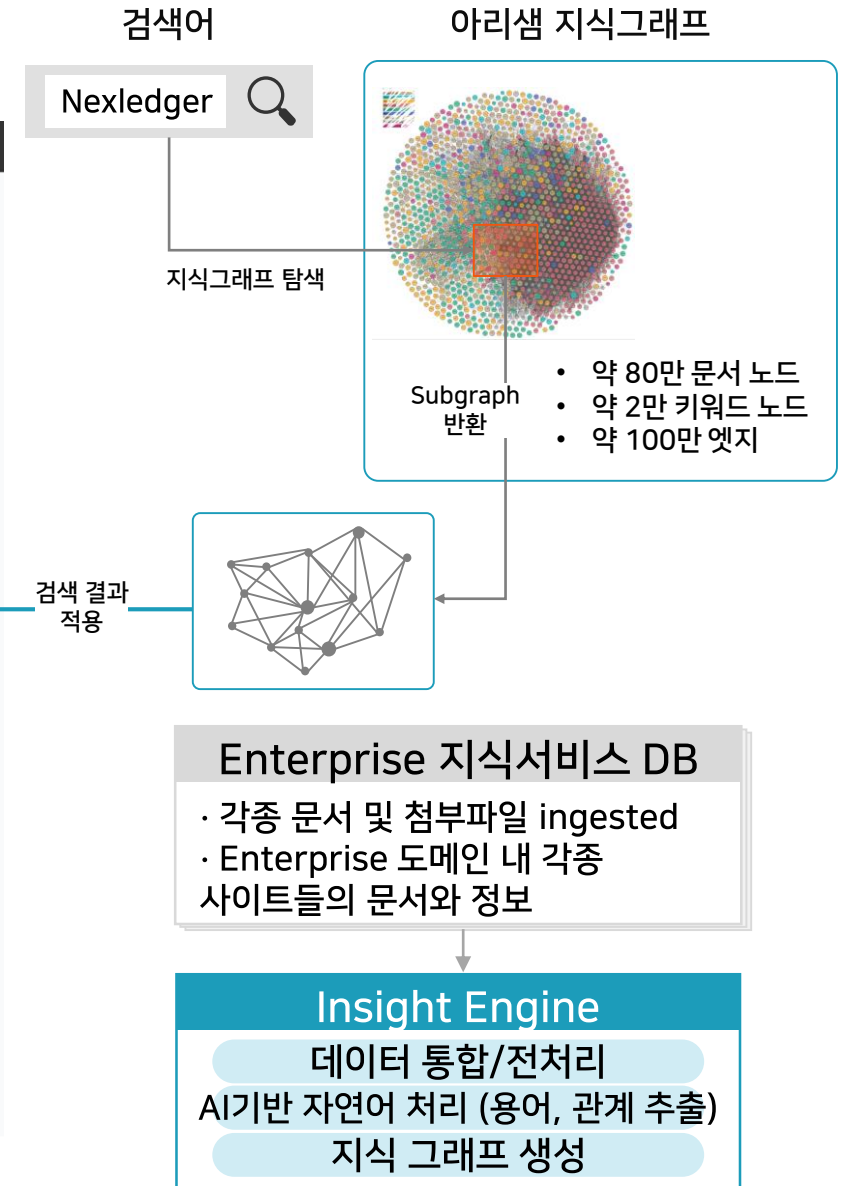
검색 결과 적용

관련 사업 정보

- * [Nexledger] - [US]
- * [Nexledger] - [한]
- * [Nexledger] - [MT]
- * [Nexledger] - [주-]
- * [Nexledger] - [한]

관련 임직원 정보

- * SCP
- * 영일대표
- * 재인PM



Case Study 4: AICR



기존 OCR과 달리 AI 기반의 이미지화된 서식과 문자를 인식해서 문자화된 정보를 자동 추출함으로 비즈니스 업무 자동화를 지원하는 삼성SDS의 딥러닝 솔루션

▶ AI 기반 문자 · 문서 구조 인식엔진

납세증명서

납세번호: 8871 289 처리기간: 국세(단, 해외이주용 10일)

납세자: 정호(정인영) 사업자등록번호: 220-88

납세자 인적사항: 주소(본원): 서울특별시

납세자의 사유: [] 미납수입 [] 해외이주 (이주번호: 제 호, 이주확인일: 년 월 일)

납세자의 납부기간: 2019년 11월 17일

납세자의 납부기간: [] 기납 (사유:)

납세자의 납부: 유예종류 유예기간 유예기간 세액 납부기간 세액 가산금

납세자의 납부: 해 당 없 은

납세자의 납부: (단위: 원)

납세자의 납부: 위약금 유예기간 세액 납부기간 세액 가산금

납세자의 납부: 해 당 없 은

납세자의 납부: (단위: 원)

국세징수법 제6조 및 같은 법 시행령 제6조에 따라 납부일 현재 위의 납세유예액, 제납지불유예액 또는 「부가가치세법」 제3조제2에 따른 수납자의 불납납세의무와 관련된 제납액을 제외하고는 다른 제납액이 없음을 증명합니다.

납세번호: 501607879892

납세부서: 민원봉사실

납세자: 2019년 10월 18일

납세처: 02-840-9223

동작세무서장

국세청

국세청의 제2호지침에 따라 발급된 것으로 90일 이내 「국세청 홈택스」(www.hometax.go.kr) 또는 모바일 홈택스 > 국세청(민원) > 납세증명서 발급하기 >에서 발급받을 수 있음. 또는 국세청의 제2호지침에 따라 발급된 것으로 확인이 가능함입니다. <중복제출된 납세증명서>의 경우는 10일 이하의 기간에 삭제될 수 있습니다.

본 증명서를 증명서(www.hometax.go.kr)에서 개인 증명서 서비스를 통해 발급한 증명서입니다.

문단

- 문자와 문자 형식을 기준으로 문자 영역에 해당하는 문단 인식

테이블

- 표, 리스트 구조 등 테이블과 셀 내용 인식

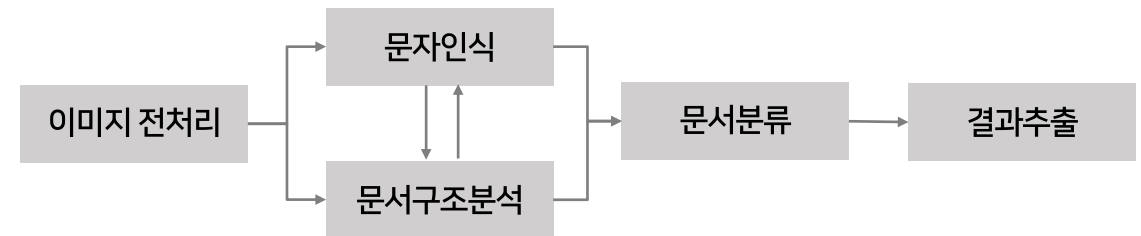
그림

- 노이즈 인식을 방지하기 위한 非 문자 영역 인식

▶ 딥러닝 기반 구조 분석

- 문자 판독 중심의 OCR과 달리, 문자와 서식, 문서 구조를 판단

▶ 핵심 프로세스 구성과 기능



[자동 보정]

- 회전, 구겨짐
- 노이즈 제거
- 디테일 강화

[문자, 서식 추론]

- 자유서식 표
- 외곽선 없는 표/셀
- 문단, 이미지

[자동 분류]

- 인보이스 (A사, B사...)
- 잔액계좌증명서 ...
- 사업자등록증 ...

[자동 변환]

- 용어 자동 치환
- 데이터 저장
- 커스텀 룰 적용

Case Study 4: AICR (cont.)

자금담당자의 전수 또는 샘플링 대사 (對査) 업무의 효율화 추진
(재무 구매 업무, 금융분야 등에 수십 건에 달하는 많은 사업 레퍼런스)

AICR 적용 대사업무 자동화

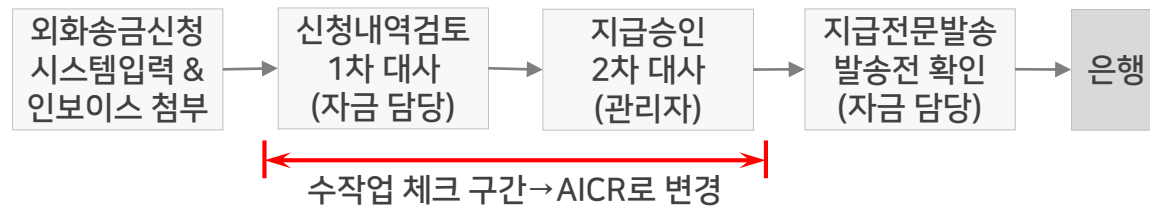
외화송금인보이스 사례

인보이스와 SAP에 입력된 외화송금 신청내역 검증

- 평균 21,000여 건 / 년 자동화
- 양식 400여 종
- 인건비 72% 절감

● 불일치
■ 일 치

업무 프로세스



Foreign Currency FB Request Review

Payment Proposal Reject Refresh Mass Change AICR I/F Attached File Download

Document date	Sele...	Result	1	2	3
2020.	<input type="checkbox"/>	⚠	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■
2020.	<input type="checkbox"/>	○	■	■	■

Bill to : GAMGUNG GDO CO LTD

SAMSUNG SDS VIETNAM COMPANY LIMITED ①

TAX INVOICE

⑦ Invoice No. : EP
⑧ Invoice Date : 03
Customer No. : T11
Due Date : 03
Customer Ref. No. : 90X
Our Ref. No. : SIC

Bank Information :
CITI BANH N.A-HANOI BRANCH
1st Floor, 17 Ngo Quyen Street, Hoan
Kiem Dist, Hanoi, Vietnam
② Account No. : 0201
④ Swift Code : CITI

시스템 : Vietnam Co. Ltd

문 서 : VIETNAM COMPANY LIMITED

Foreign Currency FB Request Review

Status	Name	Treasury	AICR
①	Beneficiary	Samsung SDS Vietnam Co., Ltd	Samsung SD
②	Account No.	02	02
③	IBAN		
④	Swift Code	CITI	CITI
⑤	Currency	USD	USD
⑥	Invoice Amount	6288.00	6288.00
⑦	Invoice Number	EP10	EP10
⑧	Invoice Date	2020	2020

Description	Period of Sales	Amount	Currency
		6,288.00	USD
		6,288.00	USD
		0.00	USD
		6,288.00	USD

⑥ ⑤

Stamp: SAMSUNG SDS VIETNAM, GIAM ĐỐC PARK SE JOON

Case Study 5: 강화학습

정답 제공이 어려운 상황에서 행동(Action)에 대한 보상(Reward)을 가지고 학습하여 최적화된 방법을 찾아내는 강화학습(Reinforcement Learning)으로 비즈니스 난제를 해결

물류 적재 및 운송, 재고 창고 관리, 자원 최적화, 생산 설비 운영, 반도체 패키지 설계 등 조합 최적화를 요구하는 대부분 NP-hard 문제에 적용 가능

물류 적재 최적화

- Allocation Problem
- 물류창고 적재 최적화 시뮬레이션에 적용

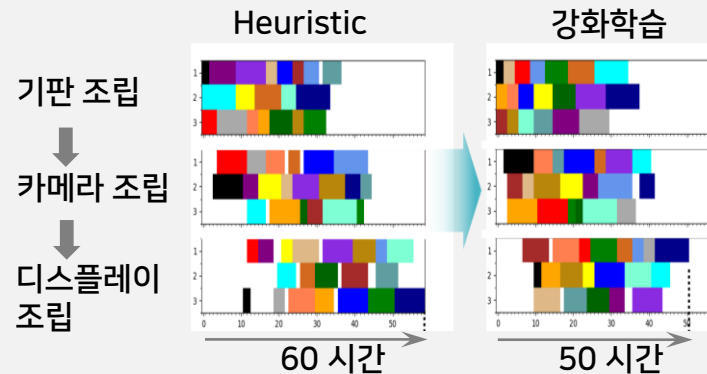
구분	컨테이너	팔레트	결과
기존 엔진 (Meta heuristics)			20%유형 개선 효과
강화학습 (SDS POMO)			

POMO: Policy Optimization with Multiple Optima

제조 공정 Scheduling

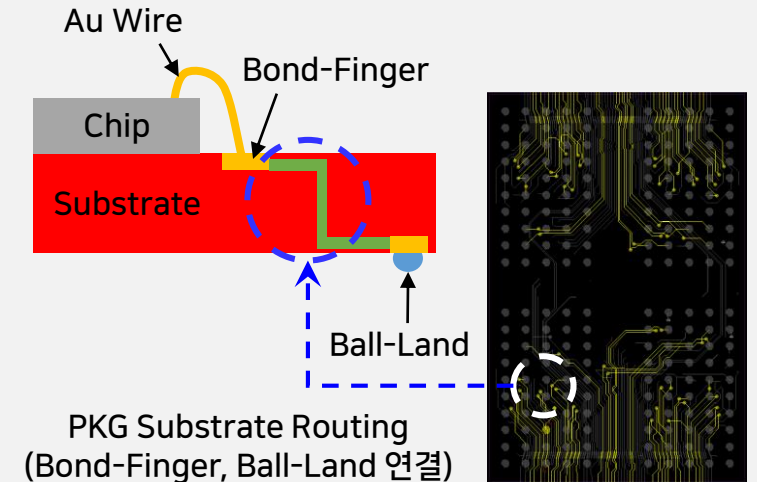
- Flexible Flow Shop Problem
- 최적 생산 scheduling을 통한 공정효율 개선

3개 프로세스, 3개 설비, 20개 작업 스케줄링



반도체 PKG Auto Routing

- Routing Optimization Problem
- 반도체 multi-layer 구조 상 연결선의 최적 배선설계



Summary

- AI aims “on-par” with human ability
- Linguistic and visual intelligence crucial for developing AI for enterprise business applications
- Thank you!