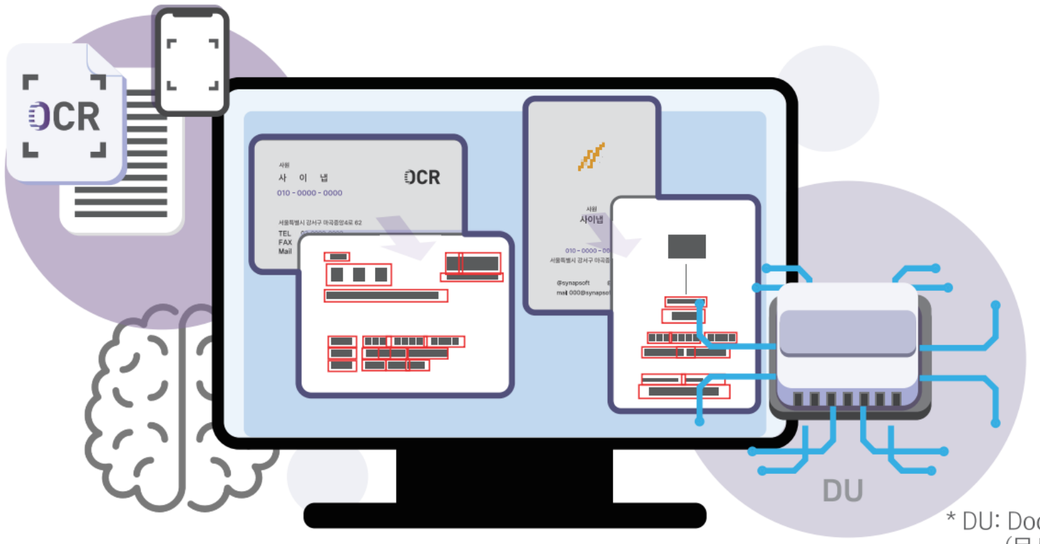
사이냅소프트

최신 AI 기술로 더 강력해진

사이냅 OCR Pro



* DU: Document Understanding
  (문서이해기술)
* KVP: Key-Value Pair(키-값 쌍)

Copyright Synapsoft Corp. All right reserved.

CONTENTS

01 솔루션 소개

02 특장점

03 주요 기능

04 품질 및 성능

05 활용 분야

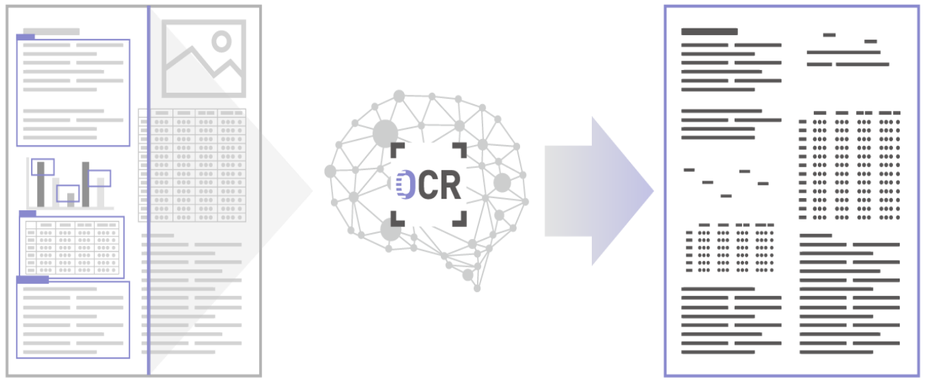
06 레퍼런스

07 회사 소개

Copyright Synapsoft Corp. All right reserved.

OCR Pro?

뛰어난 한글 인식률과 빠르고 안정적인 인식 속도, 문서 이해(DU)기술 적용으로 비정형 서식 이미지 내 핵심 정보(KVP) 추출이 탁월한 AI OCR



발전성
딥러닝을 통한 지속적인 인식률 향상

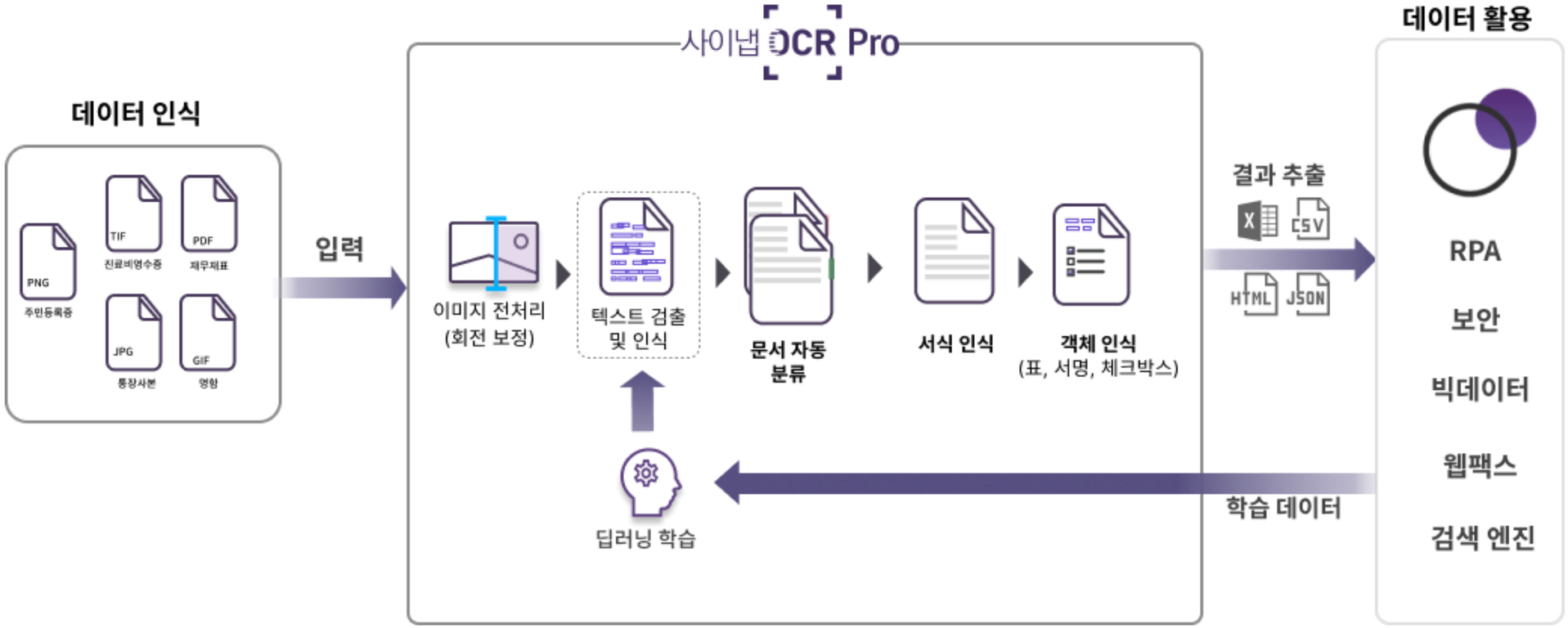
호환성
다양한 SDK, 서식 제작 도구와 학습•운영 도구 제공으로 간편하고 유연한 개발

신뢰성
다수 RPA, 검색, 보안 솔루션과의 협업으로 검증

Copyright Synapsoft Corp. All right reserved.

솔루션 흐름도

다양한 유형의 이미지에서 텍스트를 검출뿐만 아니라 문서 자동 분류, 서식 인식, 표 및 서명 등 객체 인식 지원



Copyright Synapsoft Corp. All right reserved.

OCR Pro 특장점



| 최고 수준의 한글 인식률

- 한국정보통신기술협회(TTA) 인증 공공행정문서 인식률 99.3% 달성
- 디지털 문서 처리 노하우를 바탕으로 학습 데이터를 구축·활용하여 타사 대비 높은 한글 인식률



| AI 문서 이해(DU) 기술 적용으로 비정형 서식 인식 기능 제공

- 비정형 서식을 직접 학습하고 운영할 수있는 딥러닝 모델 학습·운영 도구 제공
- 명함, 영수증, 처방전 등과 같이 발행 기관마다 서로 다른 비정형 서식도 유의미한 정보 쉽게 추출



| 템플릿 기반 정형•반정형 서식 인식 도구

- GUI 기반 직관적인 서식 템플릿 제작 도구 제공
- 정형 서식 및 기업 업무 양식 등 쉽게 등록하여 별도 학습 없이도 자동으로 서식 인식 및 문서 분류 지원



| 고객 니즈를 반영하여 개발된 SDK 제공

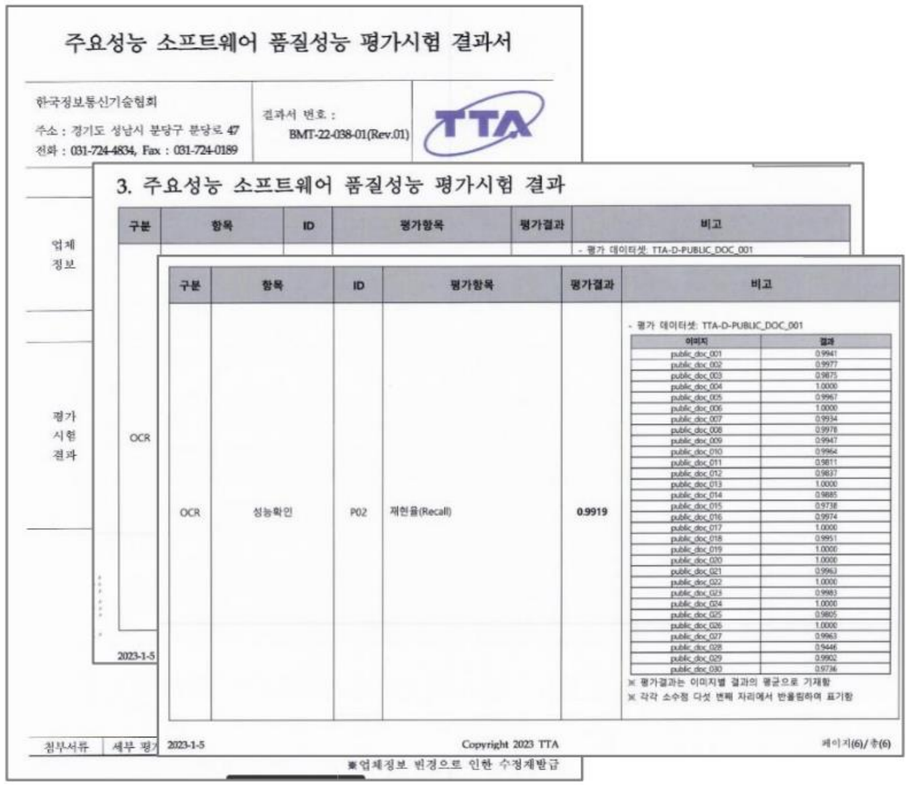
- SDK(C++, Java, Rest API, 폼페이커, KVT, OCR2PDF)를 제공하여 다양한 환경과 시스템에 유연하게 적용 가능

Copyright Synapsoft Corp. All right reserved.

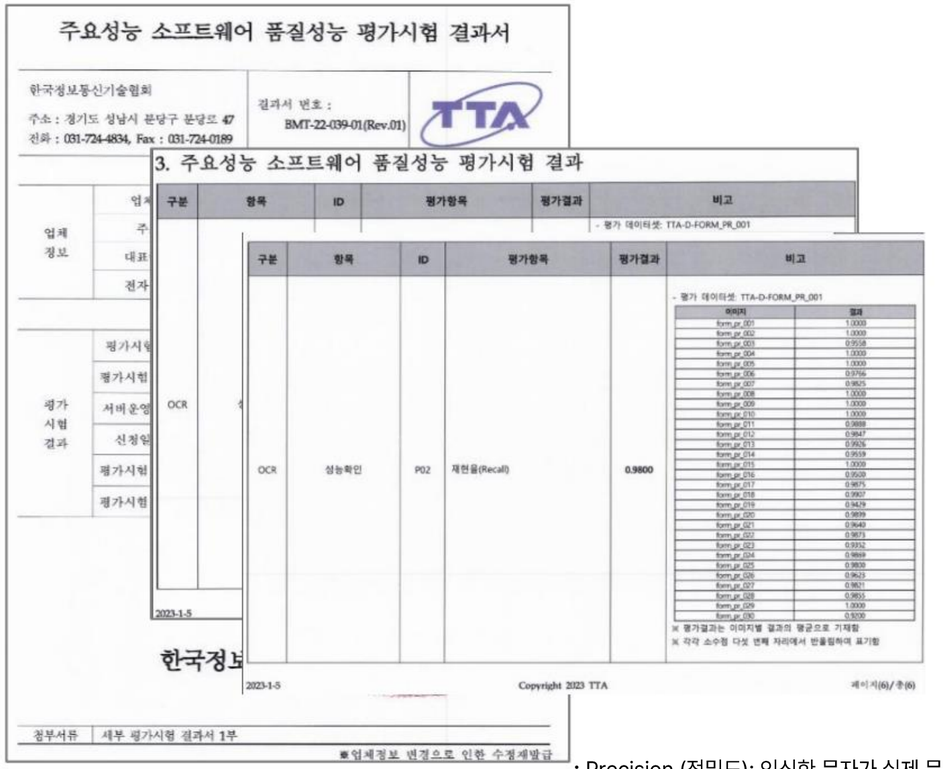
최고 수준의 한글 인식률 | 검증 받은 인식률

한국 정보통신기술협의회(TTA)에서 수행한 OCR 품질 성능 BMT 시험 수행에서 뛰어난 품질 성능을 검증받은 솔루션

항목	정밀도	재현률
공공행정문서	99.3%	99.2%



항목	정밀도	재현률
정형 서식	98%	98%



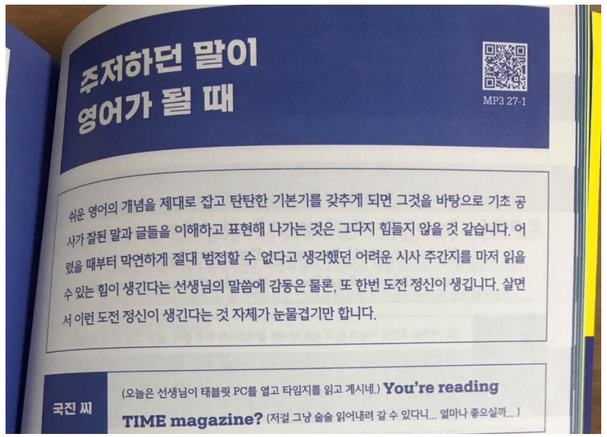
- Precision (정밀도): 인식한 문자가 실제 문서에 포함되어 있는 비율
- Recall (재현율): 실제 문서에 포함된 문자 중 인식한 문자의 비율

Copyright Synapsoft Corp. All right reserved.

최고 수준의 한글 인식률 | 실생활 이미지 인식

다양한 경로로 촬영된 실생활 이미지의 굴곡, 각도 등 자동 회전 보정 기능을 통해 높은 인식률 제공

원본 이미지

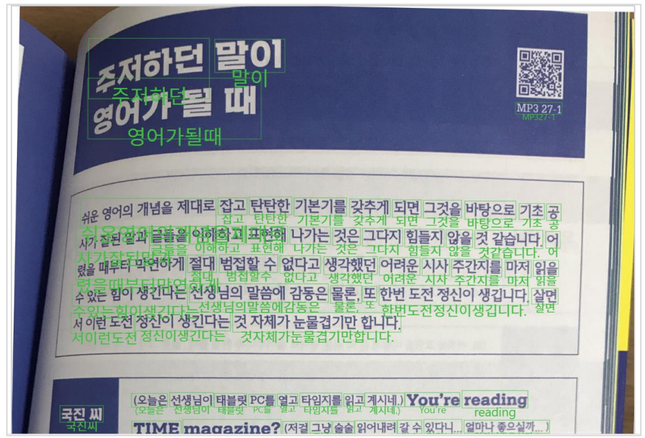


■ 이미지 굴곡



■ 이미지 회전

OCR 결과





Copyright Synapsoft Corp. All right reserved.

최고 수준의 한글 인식률 | 복잡한 배경,저품질 이미지

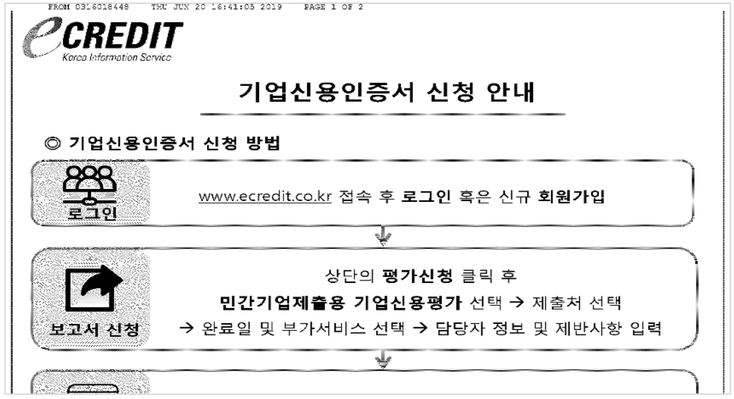
도형, 색상이 포함된 복잡한 이미지나 저해상도 이미지에 대한 높은 인식률 제공

원본 이미지

○ 사업추진 일정 계획

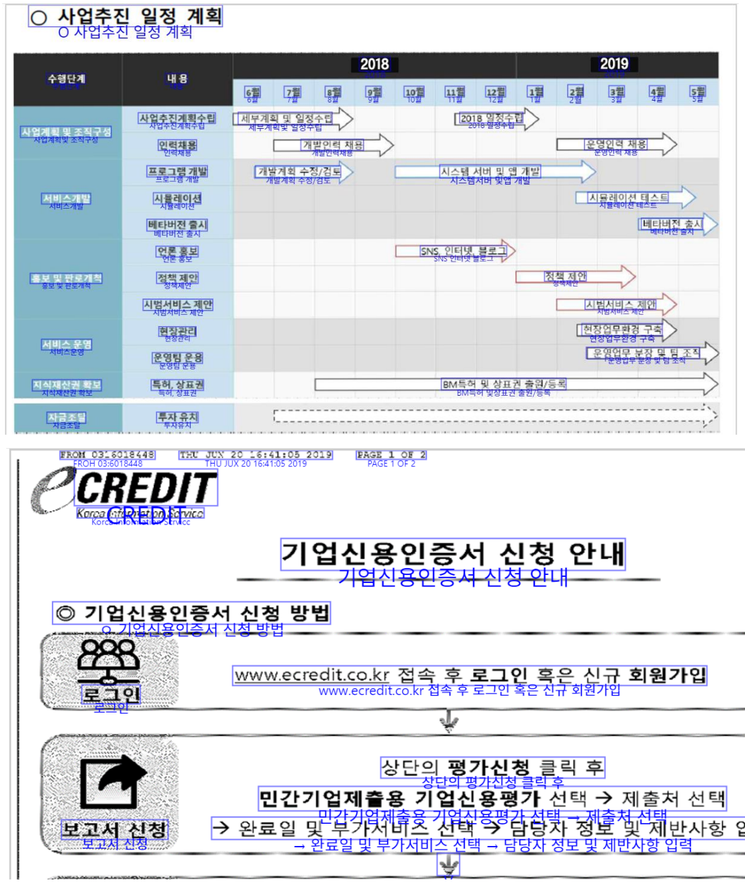


■ 도형, 색상 포함 복잡한 배경



■ 저해상도

OCR 결과

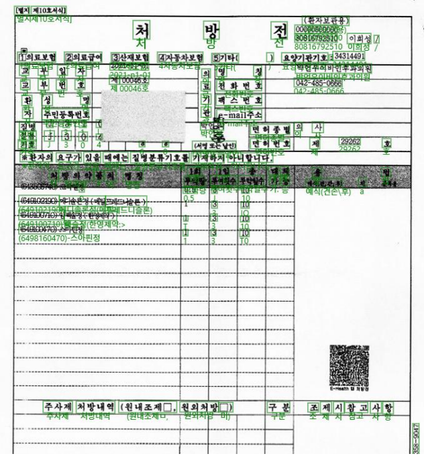


Copyright Synapsoft Corp. All right reserved.

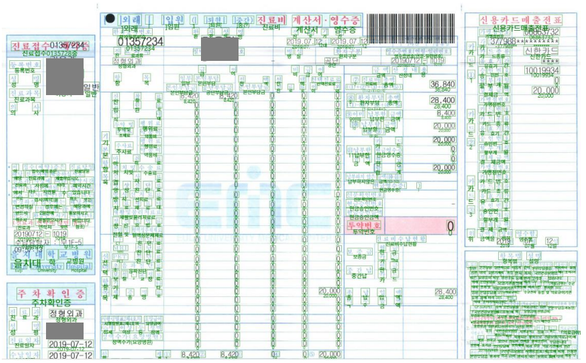
비정형 서식 인식 기능 제공 | 다양한 비정형 서식 인식

비정형 서식 학습을 통해 처방전, 통장사본, 명함, 영수증 등 다양한 유형의 비정형 서식에서도 데이터를 빠르고 정확하게 추출

보험 청구 서류



■ 처방전

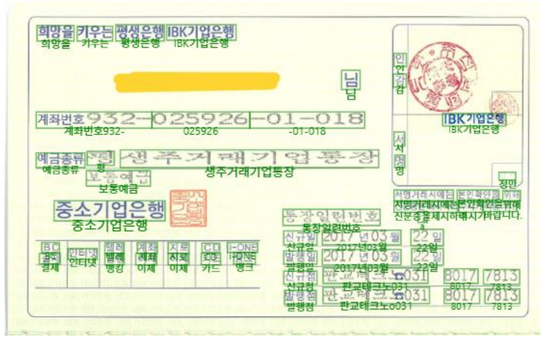


■ 진료비 영수증



■ 진단서

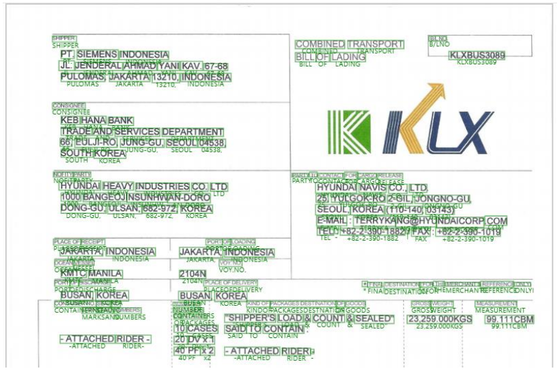
기타 행정 서류



■ 통장사본



■ 영수증



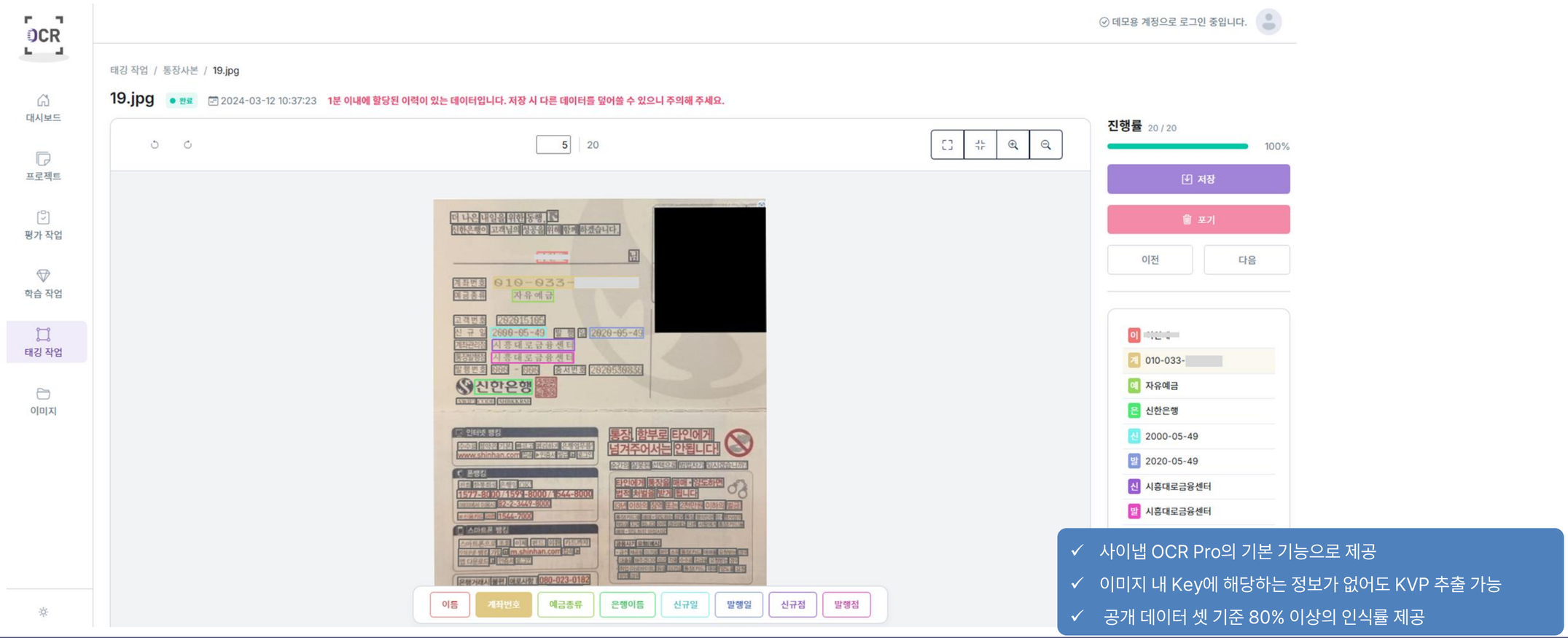
■ 송장(invoice)

Copyright Synapsoft Corp. All right reserved.

비정형 서식 인식 기능 제공 | 학습기반 서식 도구 제공 (KVT)

학습 기반 서식 도구 KVT(Key Value Trainer)는, 문서 이해 분야의 다양한 Business Task(Downstream Task) 중 SER(Semantic Entity Recognition)기술을 사용하여 학습 및 추론을 통해 비정형 이미지에서 핵심 데이터 추출

KVT 화면 예시



- ✓ 사이냅 OCR Pro의 기본 기능으로 제공
- ✓ 이미지 내 Key에 해당하는 정보가 없어도 KVP 추출 가능
- ✓ 공개 데이터 셋 기준 80% 이상의 인식률 제공

Copyright Synapsoft Corp. All right reserved.

비정형 서식 인식 기능 제공 | KVT 학습 및 배포 프로세스

KVT는 비정형 서식 인식을 위해 데이터 구축, 학습 및 평가 , 배포 및 운영의 세 단계를 거쳐 다양한 비정형 서식을 학습 및 추론을 통해 인식하며 사이냅 OCR Pro와 긴밀하게 연계하여 동작



Copyright Synapsoft Corp. All right reserved.

02
특장점
(상세)

템플릿 기반 정형·반정형 서식 인식 |위치 기반 템플릿 도구 제공

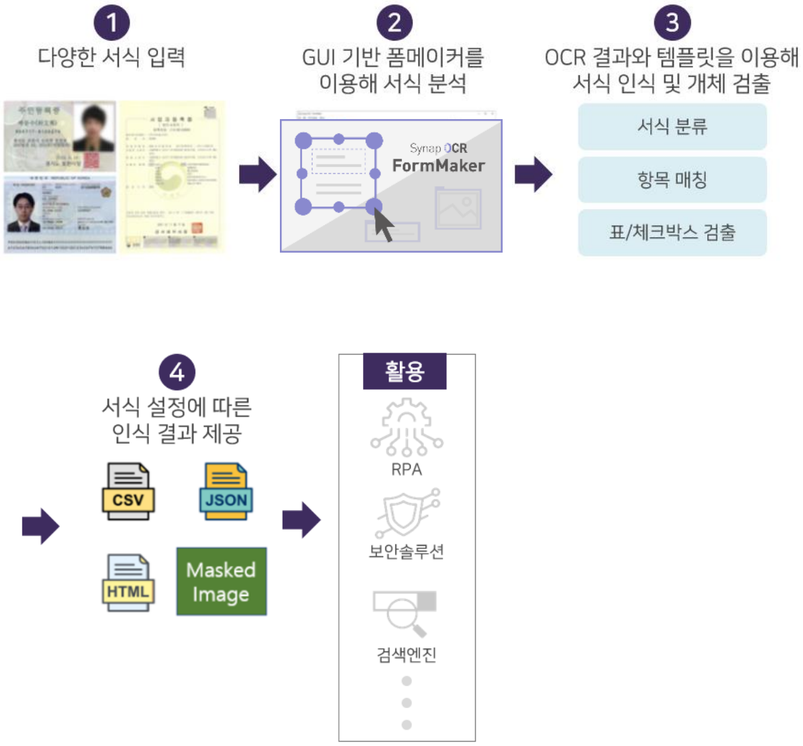
GUI 기반 직관적인 서식 템플릿 제작 도구를 제공하여 정형/반정형 문서를 자동 인식하고 템플릿 기반의 서식 인식 및 데이터 추출 기능을 통해 문서를 분류 및 데이터 추출 가능

폼메이커 화면 예시



- ✓ 템플릿 기반의 서식 인식 및 데이터 추출 제공
- ✓ 라벨링 및 핵심 Key-Value 추출
- ✓ 정의한 서식 정보에 따라 영역 마스킹 기능 제공
- ✓ 체크 박스 서식 설정 및 인식 기능 제공
- ✓ 서식 내 표 영역 인식 기능 제공

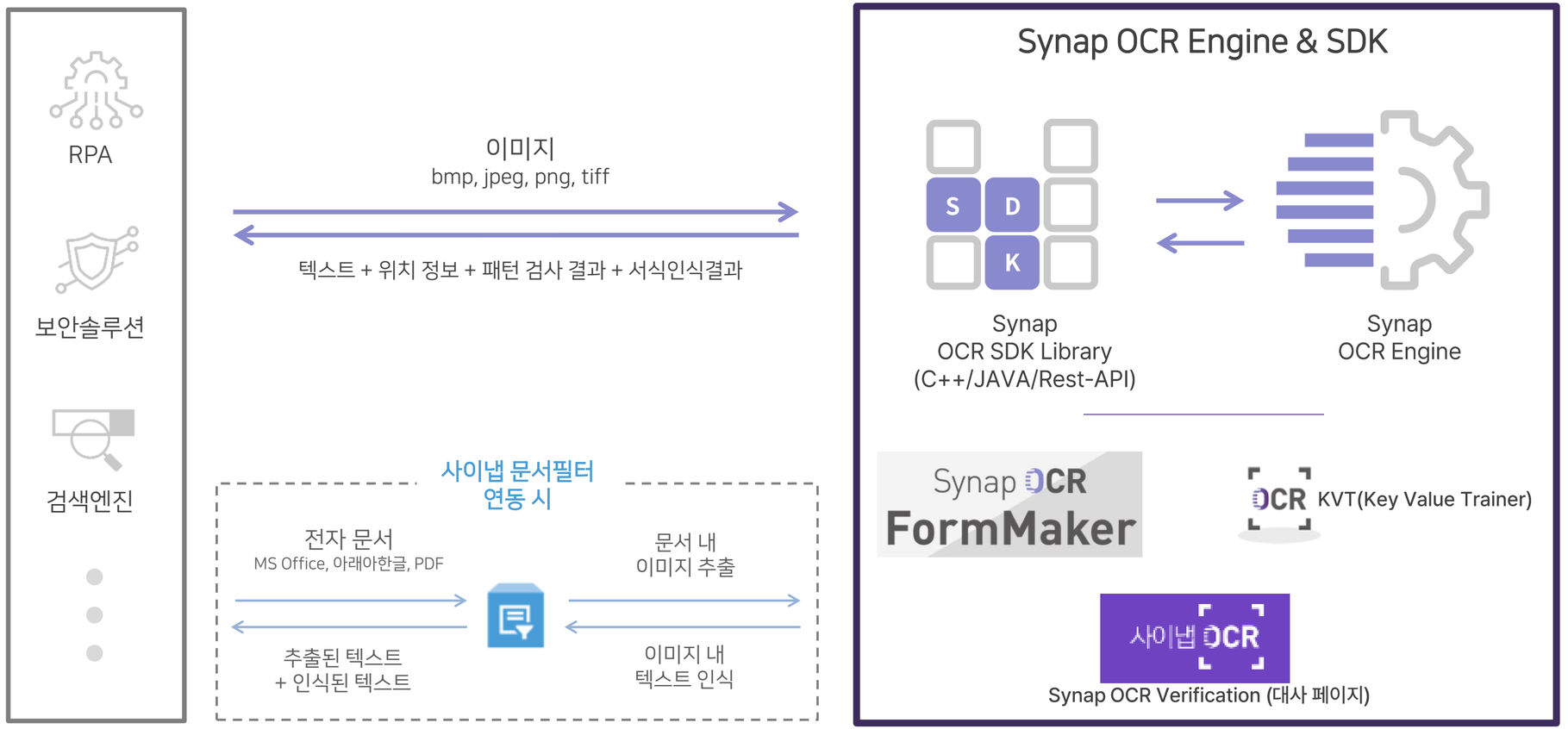
폼메이커 활용 과정



Copyright Synapsoft Corp. All right reserved.

고객 니즈를 반영하여 개발된 SDK | 솔루션 구성

SDK는 C++/JAVA 라이브러리, Rest-API, 서식 편집 도구(폼메이커), KVT, 대사 페이지가 포함되어 있으며 이를 활용하여 다양한 환경에 손쉽게 연계 구성 가능



Copyright Synapsoft Corp. All right reserved.

사이냅 OCR Pro 주요 기능

대사 페이지 제공



인식 결과의 정확성 검증 및 데이터 보정
도구인 대사 페이지 제공

표 인식 및 추출



이미지 내 표 자동 탐지 및 데이터 추출
표의 좌표 및 크기, 행/열 개수, 병합 정보 제공

다양한 서식 컴포넌트 지원



동의서, 결재서 등에 포함된 체크 박스와
서명 여부 등 서식 컴포넌트 인식 기능 제공

대형 이미지 지원



가로 혹은 세로 길이 6,000px 이상
대형 이미지에 대한 높은 인식 품질
(e.g. 쇼핑몰의 상품 설명 배너 등)

PDF 변환 및 병합



디지털 아카이빙이 가능하도록
검색 가능한 PDF로 변환 및 병합

민감정보 마스킹

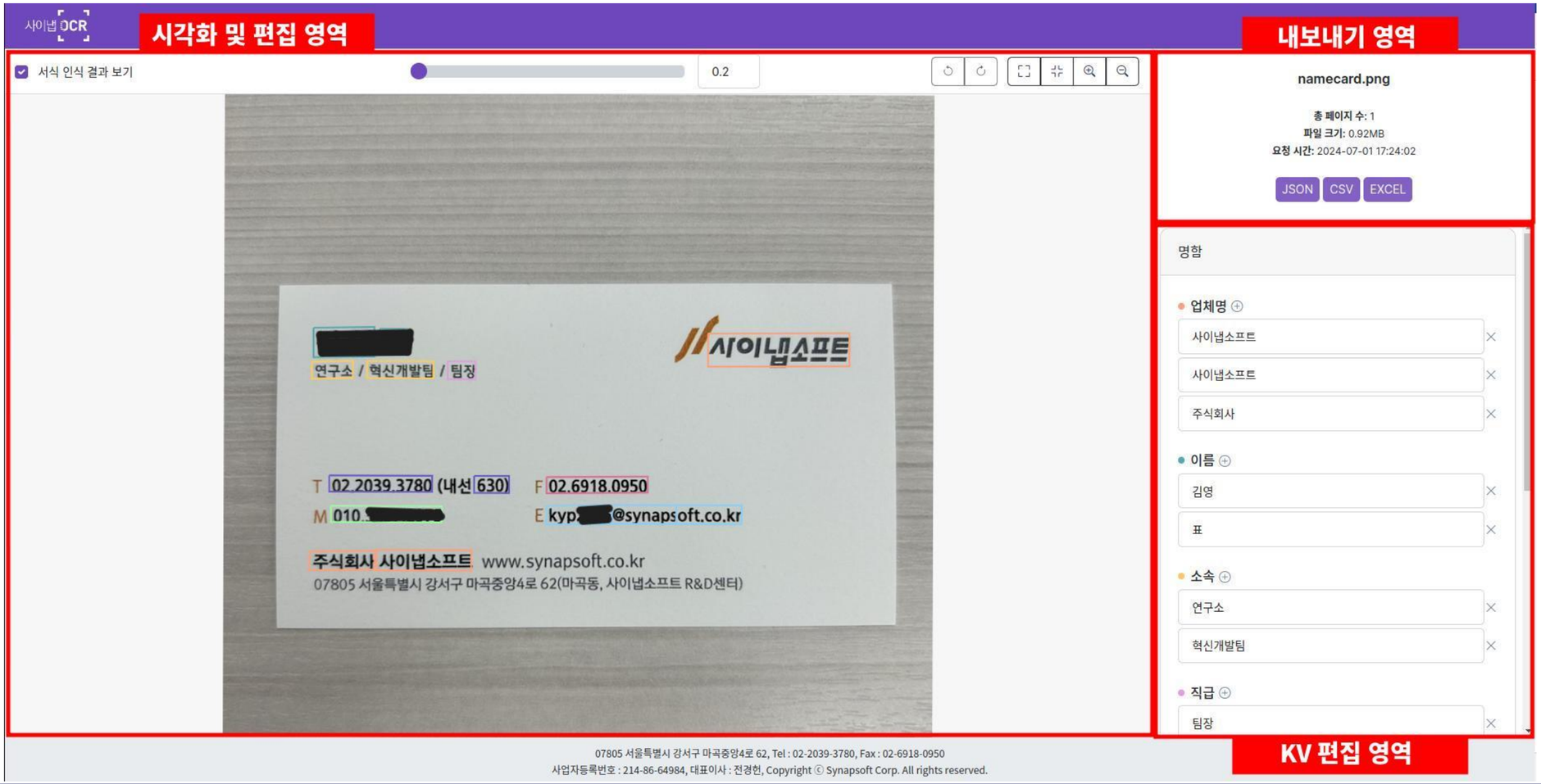


정규식 패턴 매칭 지원으로 주민번호,
여권번호 등 다양한 형태의 마스킹 가능

Copyright Synapsoft Corp. All right reserved.

대사 페이지 제공

서식 인식 결과를 시스템에 저장하기 전, 인식 결과 정확성 검증 및 데이터 보정 도구인 대사 페이지 제공



Copyright Synapsoft Corp. All right reserved.

표 인식 및 추출

이미지에 포함된 표 객체를 검출하고 데이터로 재구성할 수 있는 정보 제공

표 인식 및 추출 과정

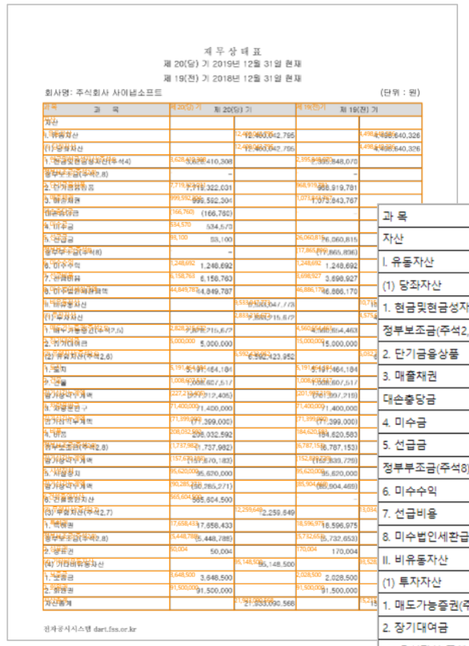


입력 이미지

전 처리 -> 표 영역 검출 -> 셀 영역 검출 -> 병합 정보 구성

- ✓ 테두리가 명확하게 있는 표에 대한 인식 기능 제공
- ✓ 표에 포함된 각 셀의 병합 정보를 정확하게 추출하여 원본과 동일한 표의 형태 정보 제공

표 인식 및 추출 결과



재 무 상 태 표
제 20(당) 기 2019년 12월 31일 현재
제 19(전) 기 2018년 12월 31일 현재

회사명: 주식회사 사이넘소프트 (단위 : 원)

과 목	제 20(당) 기		제 19(전)기	
자산				
I. 유동자산		12,400,042,795		4,498,640,326
(1) 당좌자산		12,400,042,795		4,498,640,326
1. 현금및현금성자산(주석4)	3,628,410,308		2,395,848,070	
정부보조금(주석2,8)	-		-	
2. 단기금융상품	7,719,322,031		968,919,781	
3. 매출채권	999,592,304		1,073,843,767	
대손충당금	(166,760)		-	
4. 미수금	534,570			
5. 선급금	93,100		26,060,815	
정부부조금(주석8)	-		(17,865,896)	
6. 미수수익	1,248,692		1,248,692	
7. 선급비용	6,158,763		3,698,927	
8. 미수법인세환급액	44,849,787		46,886,170	
II. 비유동자산		9,533,047,773		10,715,002,003
(1) 투자자산		2,833,215,672		4,575,654,463
1. 매도가능증권(주석2,5)	2,828,215,672		4,560,654,463	
2. 장기대여금	5,000,000		15,000,000	
(2) 유형자산(주석2,6)		6,592,423,952		6,032,784,714
1. 토지	5,191,464,184		5,191,464,184	
2. 건물	1,008,607,517		1,008,607,517	
감가상각누계액	(227,212,405)		(201,997,219)	
3. 차량운반구	71,400,000		71,400,000	
감가상각누계액	(71,399,000)		(71,399,000)	
4. 비품	208,032,592		184,620,583	
정부보조금(주석2,8)	(1,737,982)		(6,787,153)	
감가상각누계액	(157,670,183)		(152,839,729)	

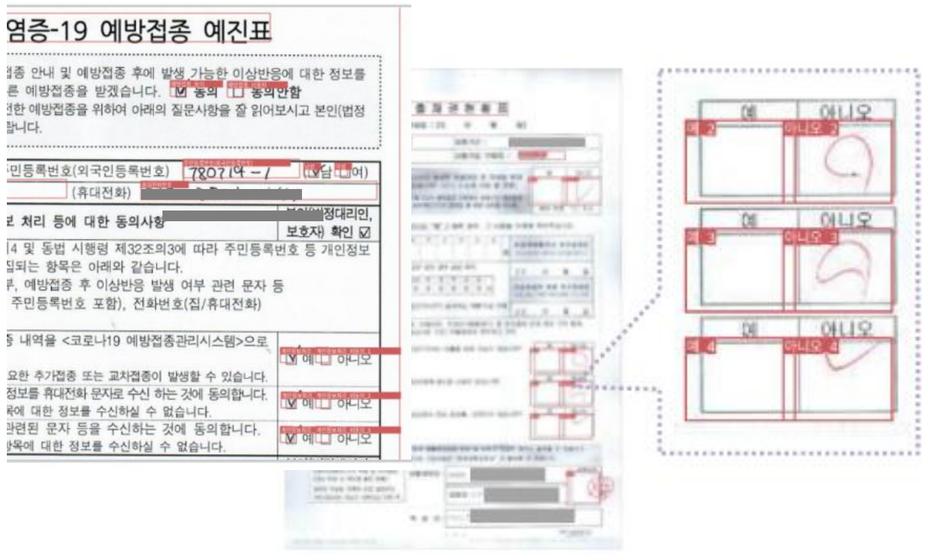
Copyright Synapsoft Corp. All right reserved.

다양한 서식 컴포넌트 지원

템플릿 도구를 활용하여 이미지에 포함된 체크 표기, 서명 여부를 확인할 수 있는 정보를 제공합니다.

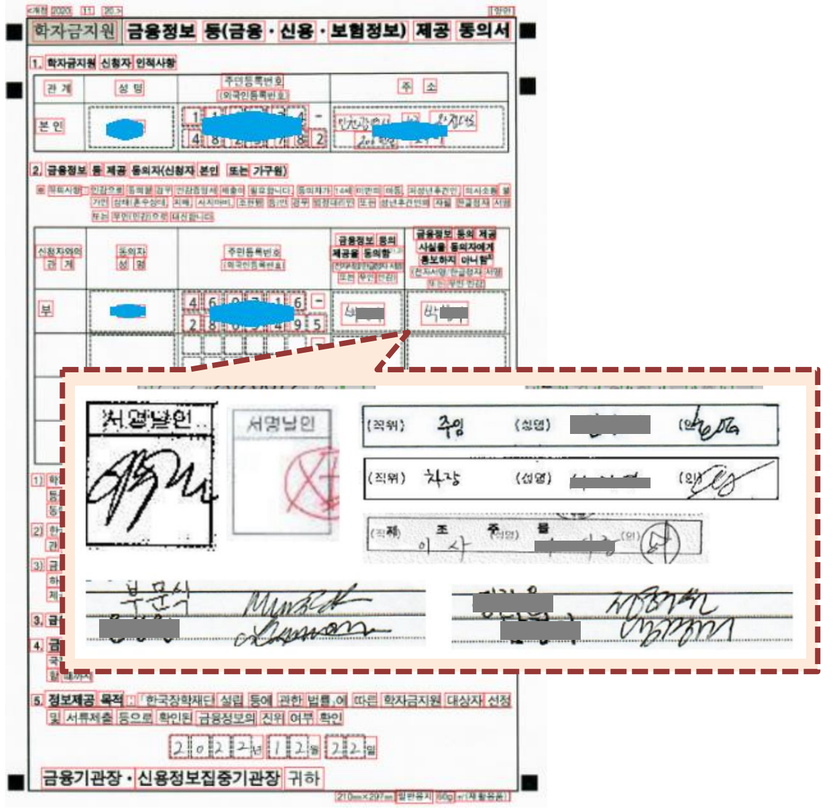
인식된 체크박스, 서명 여부 정보는 JSON 으로 제공되어 다양하게 활용 가능합니다.

체크 박스 인식



- ✓ 서식 내 수기 작성된 서명 여부 및 체크 표기 인식
- ✓ 다양한 형태로 표기된 서명 및 체크 표시를 인식하여 신청서 등의 서식 DB 구축에 활용 가능

서명 여부 인식



Copyright Synapsoft Corp. All right reserved.

대형 이미지 지원

6,000px 이상의 대형 이미지 속 글자 인식 제공, 검출된 글자의 위치 정보를 활용해 효율적인 분할 및 병합 기법을 통해 보다 정확하게 인식을 진행하며, 문자 위치 결과만 사용해 상대적으로 메모리를 효율적으로 사용하고 빠른 속도로 검출 가능

대형 이미지 인식

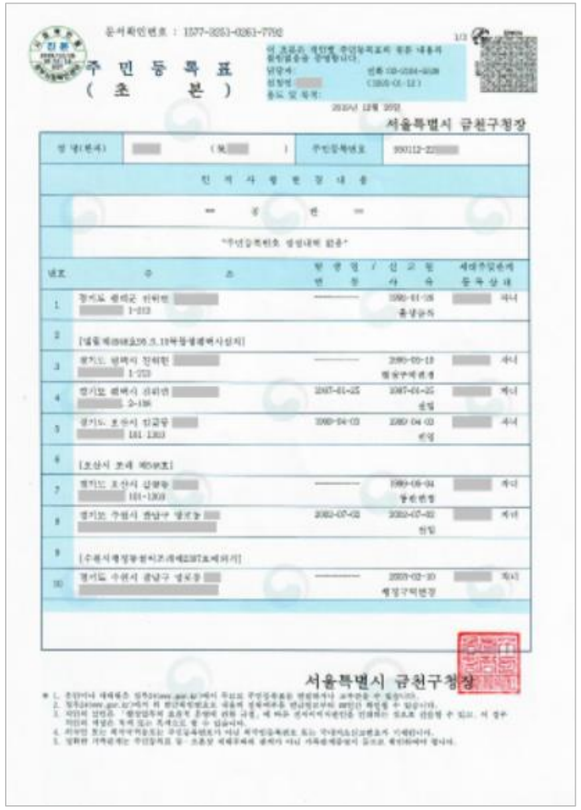


Copyright Synapsoft Corp. All right reserved.

Readable PDF 변환

검색과 편집이 가능한 Readable PDF로 변환하여 콘텐츠의 재사용성과 생산성 극대화

원본 이미지



PDF 변환 후 Drag한 모습



Copyright Synapsoft Corp. All right reserved.

민감정보 마스킹

패턴 매칭 값, 서식 도구 영역 지정을 통해 민감정보, 개인정보, 금칙어 등을 포함한 이미지를 마스킹 (또는 블러) 처리하며, 마스킹 처리된 이미지 제공

원본 이미지



마스킹 이미지



Copyright Synapsoft Corp. All right reserved.

Line Up 별 지원 기능

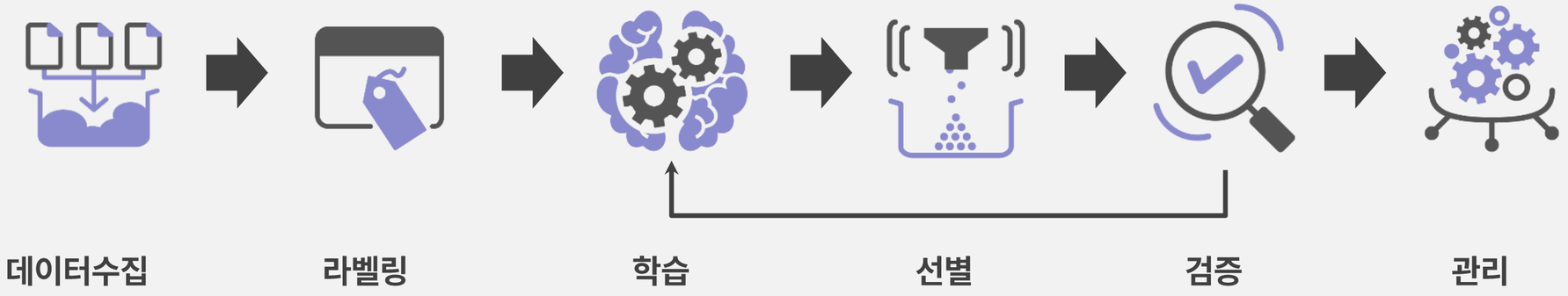
프로젝트 상황에 따른 합리적인 Line-up 구성으로 On-premise, Cloud 환경 모두 지원

기능	사이냅 OCR Pro	사이냅 OCR Lite
이미지 OCR	O	O
PDF OCR	O	O
문자 위치 기반 회전 보정	O	O
정규식 기반 마스킹	O	O
서식 인식 및 데이터 추출	O	X
서식 템플릿 작성 도구 제공 (폼 메이커)	O	
표 검출 및 인식	O	
서식 내 체크박스 및 서명 인식	O	
대형 이미지 인식	O	
PDF 변환	O	
비정형 서식 학습 도구 (KVT)	O	

Copyright Synapsoft Corp. All right reserved.

디지털 문서 처리 노하우로 고품질 OCR

인공지능 기반 디지털 문서 활용, 생성, 문서 이해 관련 토털 솔루션 전문 기업으로 검증된 문서 처리 기술을 바탕으로 대량의 OCR 학습 데이터 확보 및 정제 프로세스를 구축하여 더 높은 인식률



사이냅 OCR Pro의 데이터 확보 및 정제 프로세스

Copyright Synapsoft Corp. All right reserved.

기본 사양

하드웨어	▪ CPU: 3.0 Ghz 이상, 8코어 16쓰레드 이상 권장 ▪ Memory: 64GB 이상 권장 ▪ GPU: 10GB 이상의 NVIDIA CUDA 이상 권장 *KVT-비정형 서식 도구 활용 시 : 24 GB 이상의 NVIDIA CUDA지원 GPU 서버 별도 필요
지원 OS	▪ Linux 64bit(Ubuntu20.04,18.04,16.04), ORACLE Linux 7.6, CentOS 7.6) (KVT-비정형 서식 도구는 Linux만 지원) ▪ Windows Server 2019 64bit 이상 권장 ▪ Windows 10 64bit 이상 권장
SDK 제공 (Linux, Windows)	▪ Linux : C++ (g++ 4+), JAVA8 ▪ Windows : C++ (VC2015~VC2019, Windows VISTA 이후), JAVA8 ▪ Rest API
지원 이미지 형식	▪ Bitmap, JPEG, PNG, GIF (첫 번째 프레임) ▪ Multipage TIFF, PDF
인식 가능 문자	▪ 한글, 영문, 숫자, 생활한자, 일부 ASCII 영역의 문장 부호

Copyright Synapsoft Corp. All right reserved.

품질 수치

개별 이미지 별 문자 기준 – v2022.01

	문서 형 이미지		실 생활 이미지	
구분	특수 기호 미 포함	특수 기호 포함	특수 기호 미 포함	특수 기호 포함
Accuracy	0.8845	0.8648	0.8039	0.7925
Precision	0.8975	0.8936	0.8602	0.8513
Recall	0.9214	0.9042	0.9022	0.8974
F1	**0.9042**	**0.8935**	**0.8660**	**0.8584**

전체 기준 측정 – v2022.01

	문서 형 이미지		실 생활 이미지	
구분	특수 기호 미 포함	특수 기호 포함	특수 기호 미 포함	특수 기호 포함
Accuracy	0.9564	0.9340	0.7764	0.7647
Precision	0.9649	0.9599	0.8520	0.8419
Recall	0.9909	0.9719	0.8974	0.8929
F1	**0.9777**	**0.9658**	**0.8741**	**0.8666**

- **Accuracy (정확도)**: 전체 문자의 개수 중에 정확히 인식한 문자의 개수 비율 (TP / TN) / (TP + FN + FP + TN)
- **Precision (정밀도)**: 인식한 문자가 실제 문서에 포함되어 있는 비율 TP / (TP + FP)
- **Recall (재현율)**: 실제 문서에 포함된 문자 중 인식한 문자의 비율 TP / (TP + FN)
- **F1 (조화 평균)**: 정밀도와 재현율의 조화 평균 = 2 x (precision x recall) / (precision + recall)

- **TP (True Positive)**: 인식한 문자가 실제 문서에 포함되어 있는 개수
- **FP (False Positive)**: 인식한 문자가 실제 문서에 포함되어 있지 않은 개수
- **FN (False Negative)**: 문서에 포함되어 있으나 인식하지 못 한 문자의 개수

*품질 수치는 테스트 문서에 따라 차이가 발생할 수 있음 – 문서 형 이미지 599 장. 실생활 이미지 435 장

Copyright Synapsoft Corp. All right reserved.

성능 수치

GPU를 활용한 빠른 문자 인식과 높은 품질의 성능 제공

- Ngrinder 를 사용하여 가상의 사용자를 통해 동시 부하 발생 상황 테스트 수행
- 테스트 이미지 세트
  - 전체: 2,660 개
  - PNG: 2,550 개, JPG: 92 개, TIFF: 18개 (34페이지)
  - 10M 이상: 7개, 1M 이상: 315개, 1M 이하: 2,338개, 평균: 580kb

V2022.01, 테스트 환경 : CPU : Intel(R) Core(TM) i7-9700 CPU @ 3.00GHz, Memory : 32GB, OS : CentOS 7.5 64bit

GPU	GPU 개수	동시 요청 (부하 사용자 수)	초당 처리 건 (TPS)	평균 응답 시 (ms)	에러율(%)
RTX 2080TI 11GB	1	26	7	3540	0%
	2	37	10	3400	0%
Quadro RTX 4000 8GB	2	19	5	3330	0%

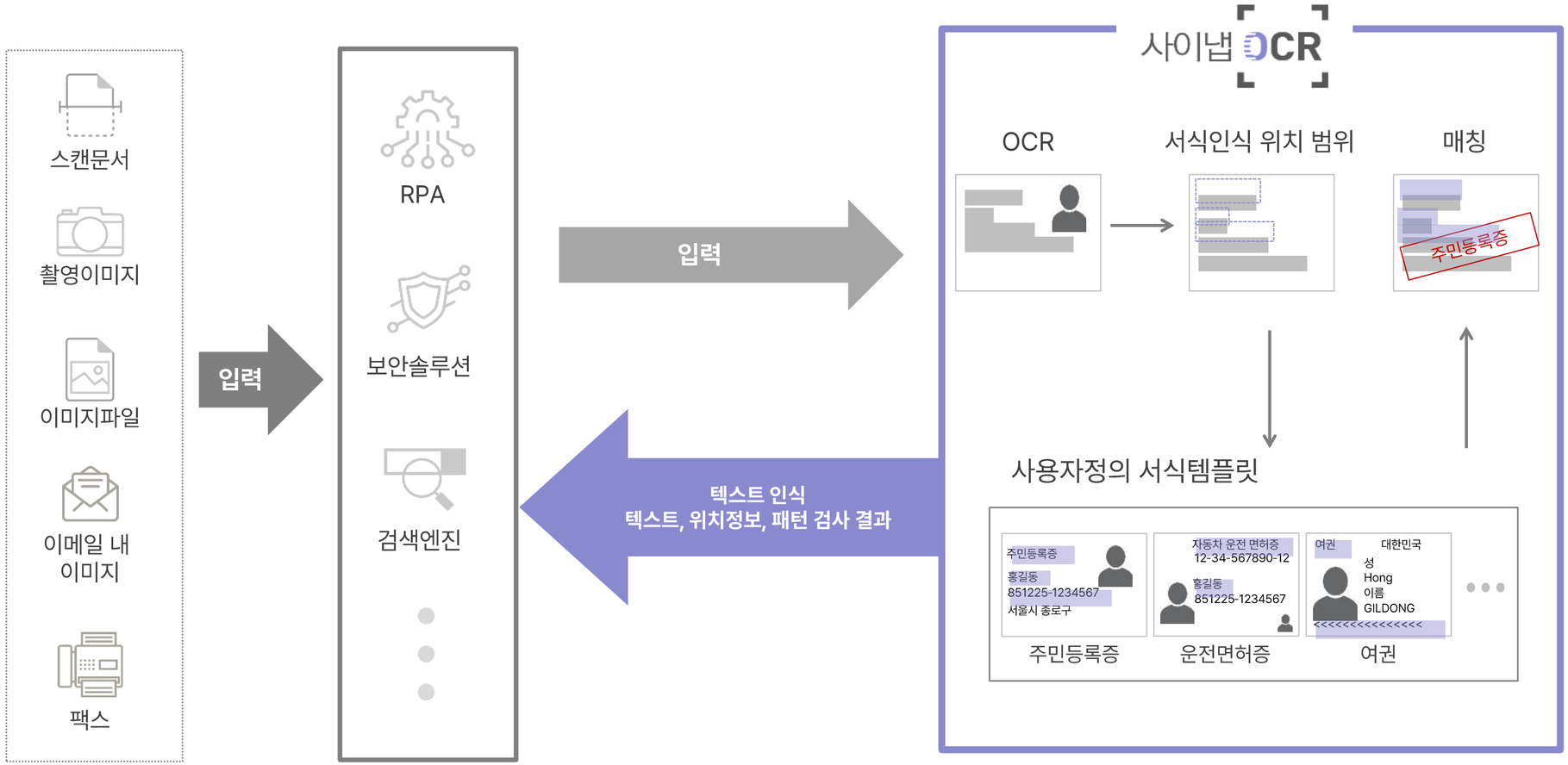
(성능 수치는 테스트 환경이나 데이터에 따라 차이가 발생할 수 있습니다.)

- TPS: 부하 상황 시 초당 처리 가능한 Transaction의 개수 (1 Transaction = 이미지 한 장의 OCR 수행 요청 후 응답)
- MTT(Mean Test Time): Transaction당 평균 응답시간
- ERR(Error Rate): 요청 별 에러율

Copyright Synapsoft Corp. All right reserved.

서식 인식을 이용한 데이터 추출 및 마스킹

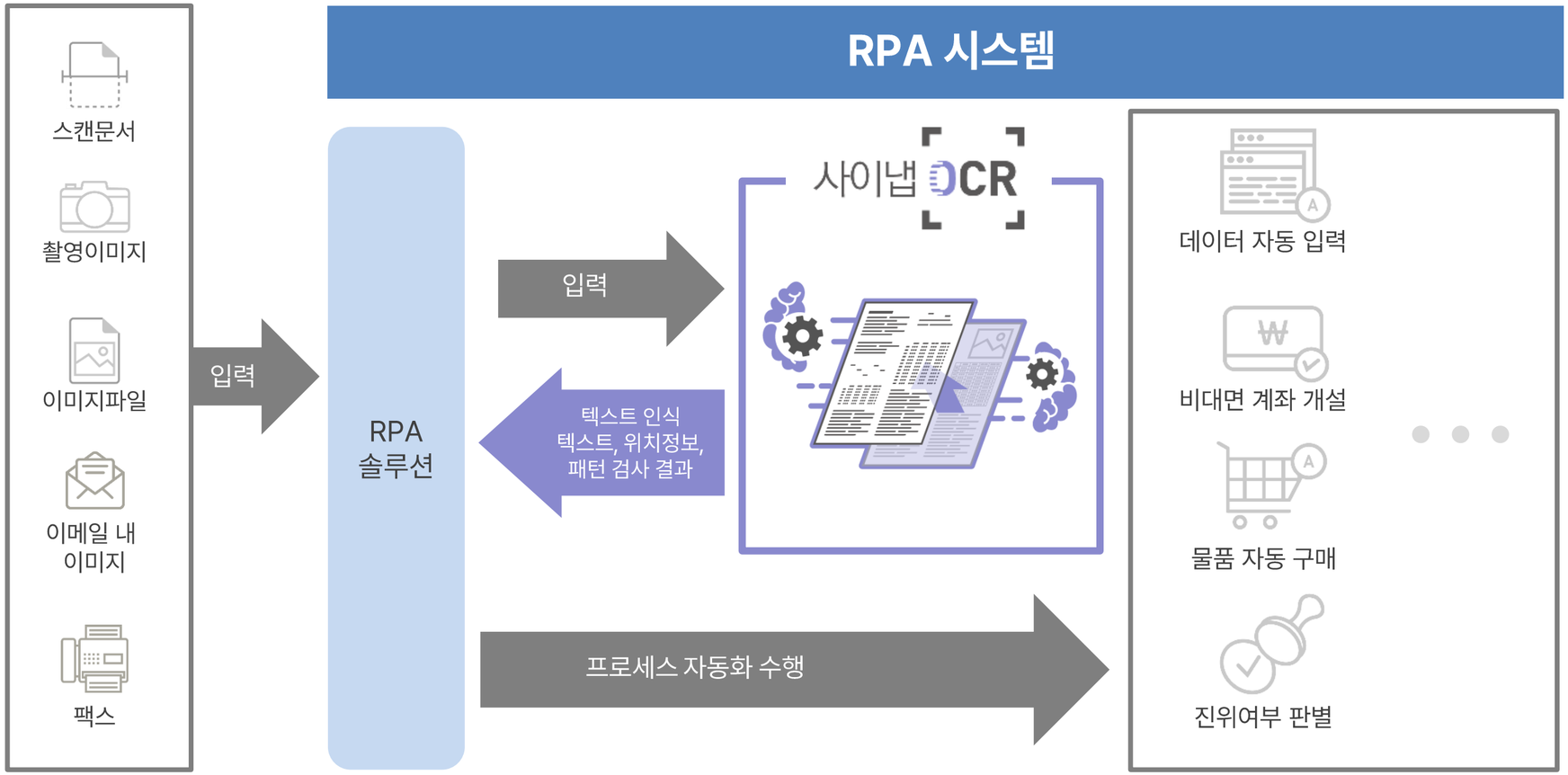
비정형 이미지 데이터를 서식 정보와 결합하여 정형화된 데이터를 다양한 시스템에서 활용 가능



Copyright Synapsoft Corp. All right reserved.

RPA (Robotic Process Automation)

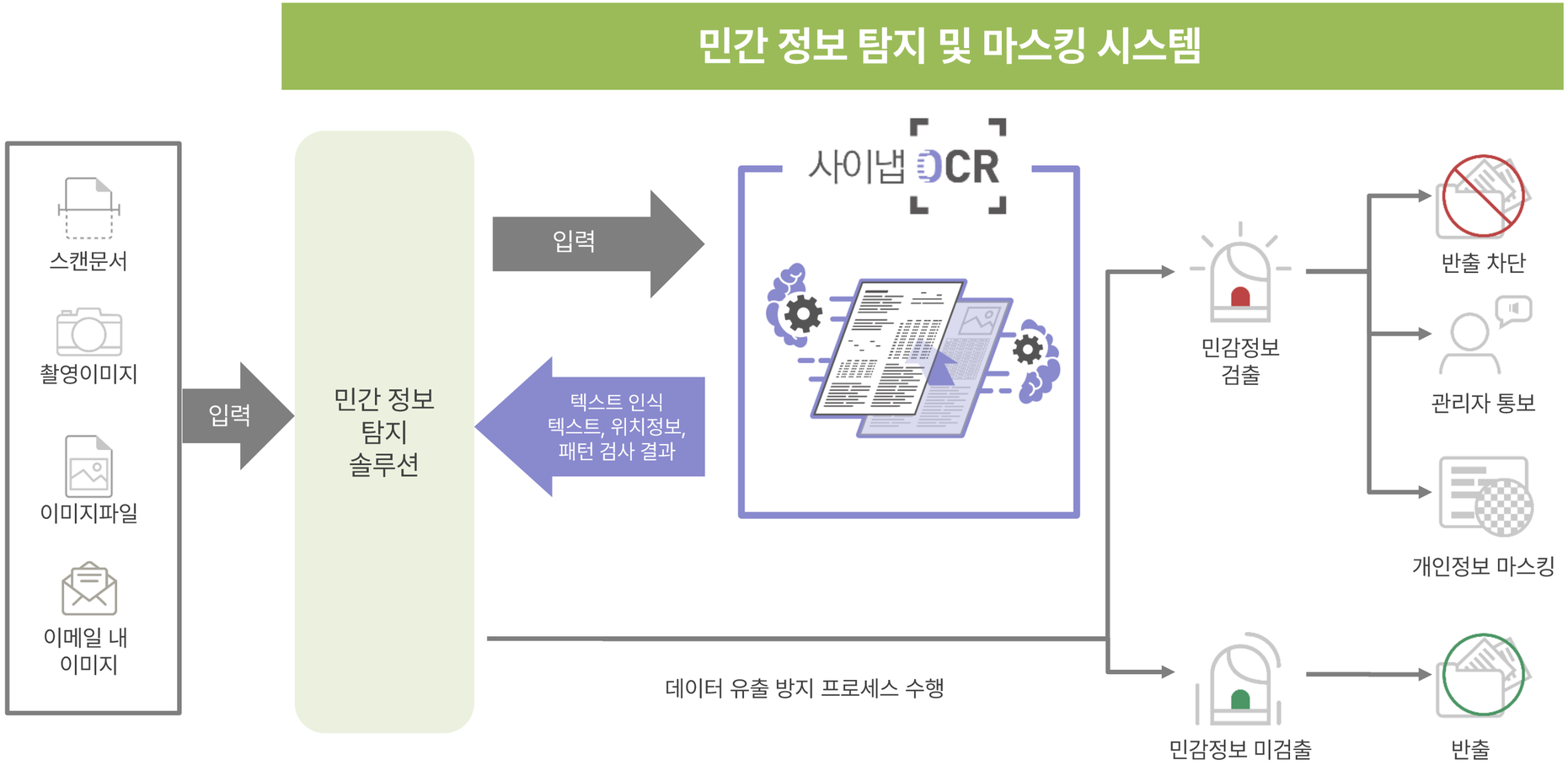
인식된 텍스트 정보는 RPA 시스템과 연동하여 다양하게 활용 가능



Copyright Synapsoft Corp. All right reserved.

민간 정보 탐지 및 마스킹 시스템

이미지, 스캔 문서에서 패턴, 키워드를 활용하여 민감 정보(주민번호, 기업 정보 등) 검출
인식된 텍스트 정보를 활용하여 솔루션에서 정보 유출 사전에 차단 가능 (자동 마스킹 처리)



Copyright Synapsoft Corp. All right reserved.

주요 도입 사례(1)

Kbank make money

회사명 : 케이뱅크

웹사이트 : www.kbanknow.com

임직원 : 403 명

매출액 : 2,878억원

*** 주요 도전과제:**

- 다양한 환경의 정형 이미지 인식

: Target 수치 90% 인식률

* 주관사, 파트너사

kt ds UiPath™ TDG

사업 개요

- 사업명 : 케이뱅크 RPA / AI OCR 기반 계정계 업무 자동화 프로젝트
- 기간 : 2021.09 - 2022.02 (6개월)
- 주관사 : KT DS (DX솔루션사업팀)
- 파트너사 : Uipath korea, TDG

핵심 과제

- 여신심사 Fax문서 처리 외 주요 다양한 정형 이미지 서식인식 :
  주민등록증,운전면허증,주민등록초본,건강보험자격득실확인서,관리비 고지서

- Uipath RPA 및 DU 연계(비정형 이미지)

구현 사항

- Form maker를 통한 정형 이미지 서식 등록 분류 및 핵심 Key Value 추출

No.	과제명	인식항목
1	주민등록증	이름, 생년월일, 발급일자
2	운전면허증	이름, 운전면허번호, 생년월일, 암호일련번호
3	주민등록초본	문서확인번호, 이름
4	건강보험 자격득실확인서	발급번호, 이름, 생년월일, 가입자구분, 사업장명, 자격취득일, 자격상실일
5	관리비고지서	주소, 고지일



Copyright Synapsoft Corp. All right reserved.

주요 도입 사례(2)

NH저축은행

회사명 :NH저축은행

웹사이트 :
www.nhsavingsbank.co.kr

임직원 : 148 명

매출액 : 1,732억원

*** 주요 도전 과제:**

- RPA와 결합하여 주민등록초본의 특정 정보 추출

사업 개요

- 사업명 : NH저축은행 RPA / AI OCR 기반 청구 자료 입·출력 자동화
- 파트너사 : BATEM (RPA)

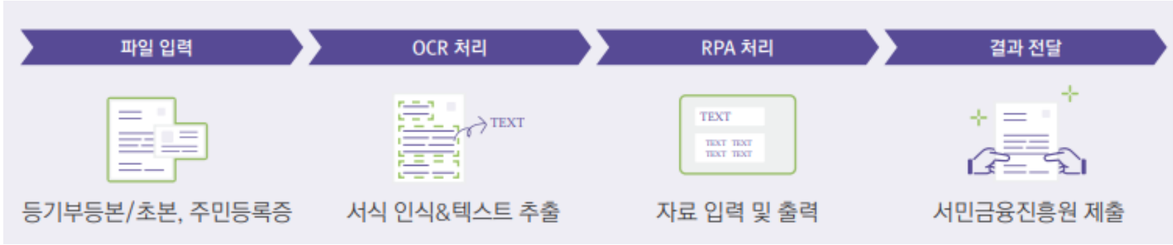
핵심 과제

- 60일 이상 연체 고객의 연체금을 서민금융진흥원에 청구하기 위한 자료 준비 업무에 OCR Pro 적용 : 주민등록증, 등기부등본/초본

- BATEM(비에이템) RPA 연계

구현 사항

- Form maker를 통한 주민등록 초본 서식 등록 및 핵심 Key Value 추출
- 4명의 담당자가 주 2회 400분 동안 수작업으로 등기부등본 등 서류 준비하는 시간을 RPA와 OCR Pro가 대신하여 업무 시간 절감



Copyright Synapsoft Corp. All right reserved.

기업 / 공공 등 검증된 AI OCR (150+)



Copyright Synapsoft Corp. All right reserved.

사이냅소프트

인공지능 디지털 문서 SaaS 전문 기업

24여 년간 개발, 축적해 온 문서 처리 원천 기술 보유

인공지능 기반 디지털 문서 활용, 생성,
문서 이해 관한 토털 솔루션 공급

85% 중앙행정부 및 7000+ 기업에 문서뷰어 공급

검색 및 보안 솔루션 분야 문서필터 시장 점유율 95%

고품질 AI OCR 사이냅 OCR Pro,
국내 최다 레퍼런스 (100+)

디지털 문서 생성

사이냅 오피스

클라우드 기반 웹오피스 스위트

Synap Word Synap Slide
Synap Cell Synap Form

사이냅 에디터

오피스를 품은 HTML5
웹에디터

디지털 문서 활용

사이냅 문서 뷰어

다운로드 없는 문서 바로보기

사이냅 문서필터

고속 문서 포맷 분석 및 텍스트
추출

사이냅 PDFocus

쉽고 안전한 PDF
생성/변환/병합

디지털 문서 이해

사이냅OCR

빠르고 정확한 딥러닝 기반
고품질 AI OCR



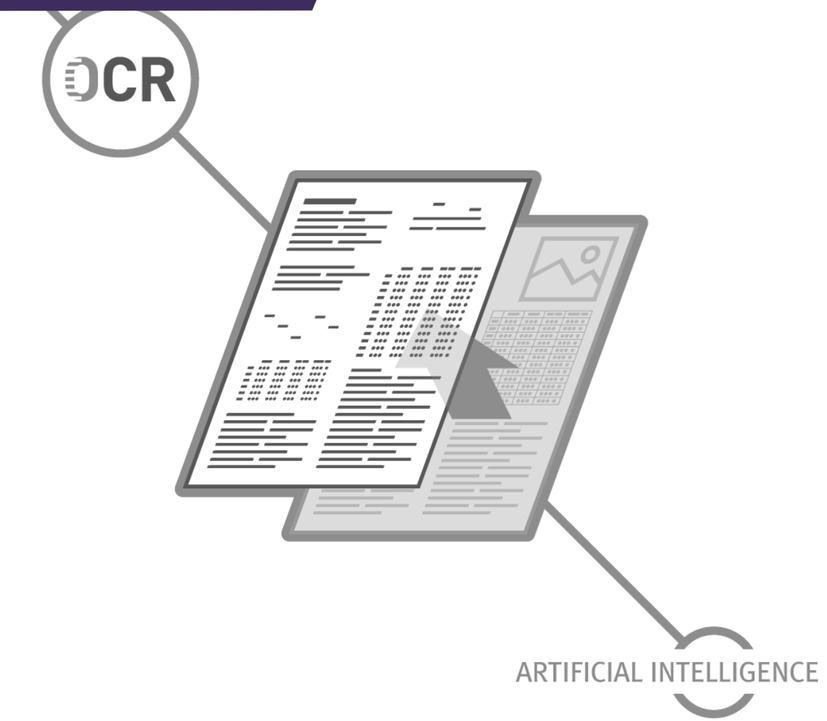
사이냅DU LLM

Synap DU

대규모 언어모델 기반
문서이해 플랫폼

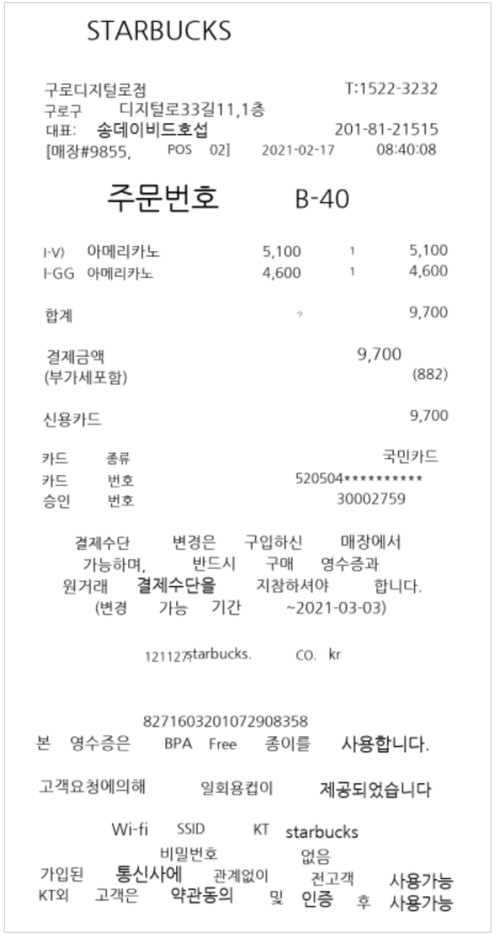
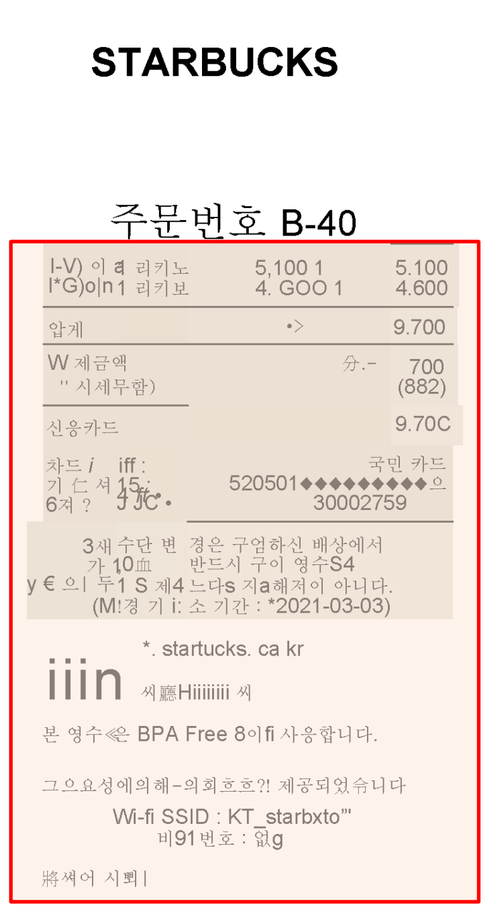
Copyright Synapsoft Corp. All right reserved.

Appendix



Copyright Synapsoft Corp. All right reserved.

영수증 (금액 단위 구분자) 인식 비교

원본	사이냅 OCR Pro 인식 결과	A사 인식결과
		
		오인식과 오검출

Copyright Synapsoft Corp. All right reserved.

Appendix

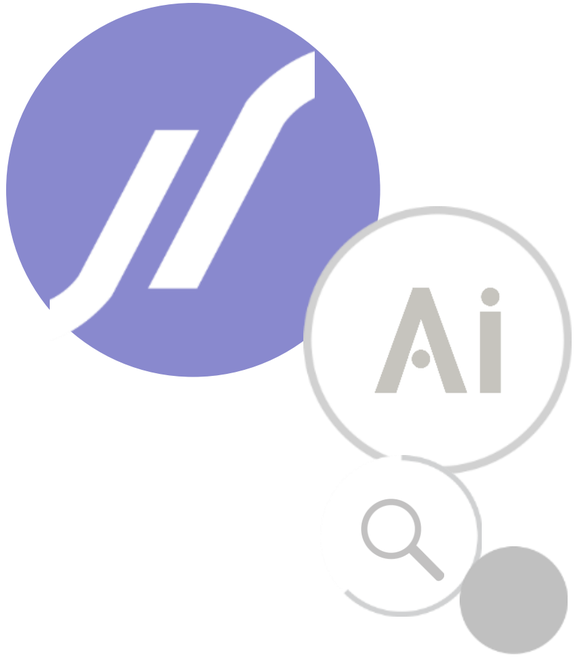
CAPCHA

원본	사이냅 OCR 인식 결과
	20643 46799 50513 67965 94563 35772 27109 85224 43584 68256 80134 98338 60733 96345 04547 11047 02774 99582 58590 70161 99580 590B 30439 13719 43332 57839 19359 65300 47467 74967
A사 인식결과	T솔루션 인식 결과
인식 실패	인식 실패

Copyright Synapsoft Corp. All right reserved.

빠르고 정확한 고품질
AI OCR

www.synapsoft.co.kr/ocr



제품 문의

솔루션사업팀

02-2039-3781 | sales@synapsoft.co.kr

서울특별시 강서구 마곡중앙4로 62 | www.synapsoft.co.kr

Youtube에서
사이냅OCR Pro를
만나보세요.

Copyright Synapsoft Corp. All right reserved.