사이냅 도큐애널라이저
SYNAP DOCUANALYZER

복잡한 문서 구조를 쉽게 **분석**하고 **가치 있는 자산**으로 만드는 최적의 시작!

©2024. SYNAPSOFT. All rights reserved.

CONTENTS

Ⅰ. 배경 및 필요성

01. 시장 배경
02. Why DocuAnalyzer?

Ⅱ. 솔루션 소개

01. 솔루션 개요
02. 주요 기능
03. 특장점
04. 시스템 기본 사양
05. 기대효과
06. 처리 성능

Ⅲ. 적용 분야

01. 적용 분야
02. 적용 사례

Ⅵ. 회사 소개

©2024. SYNAPSOFT. All rights reserved.

synapsoft

인공지능 대전환(AX)시대, LLM 활용의 품질향상을 위한 RAG의 필요성 증대

01

비정형 문서 폭발적 증가

디지털 문서, 이메일, 게시글,
이미지 등



02

정보 검색과 활용의 중요성

빠른 시장 기회 포착,
데이터 기반의 의사결정



03

디지털 자산화 필요성 증대

기업 혁신 촉진을 위한
비정형 데이터 분석



*LLM (Large language model, LLM) : 거대 언어모델
*RAG(Retrieval-Augmented Generation) : 검색 증강 생성

©2024. SYNAPSOFT. All rights reserved.

LLM 구축을 위한 방대한 비정형 문서로부터 정형 데이터 확보에 대한 고민

01

데이터 구조화의 어려움

-기업 데이터, 비정형 문서가 대다수

-문서내 표, 이미지 등 시각적 정보 포함

-데이터 일관성 유지의 한계

02

정확성 및 신뢰성 문제

-오류, 중복, 불완전 정보 포함

-잘못된 정보로 인한 결과물 품질 저하

-검증/정제 과정 막대한 리소스 투입

03

복잡한 문서 구조 처리 한계

-문서 내 표, 이미지 등
비(非)텍스트 요소 인식 부정확

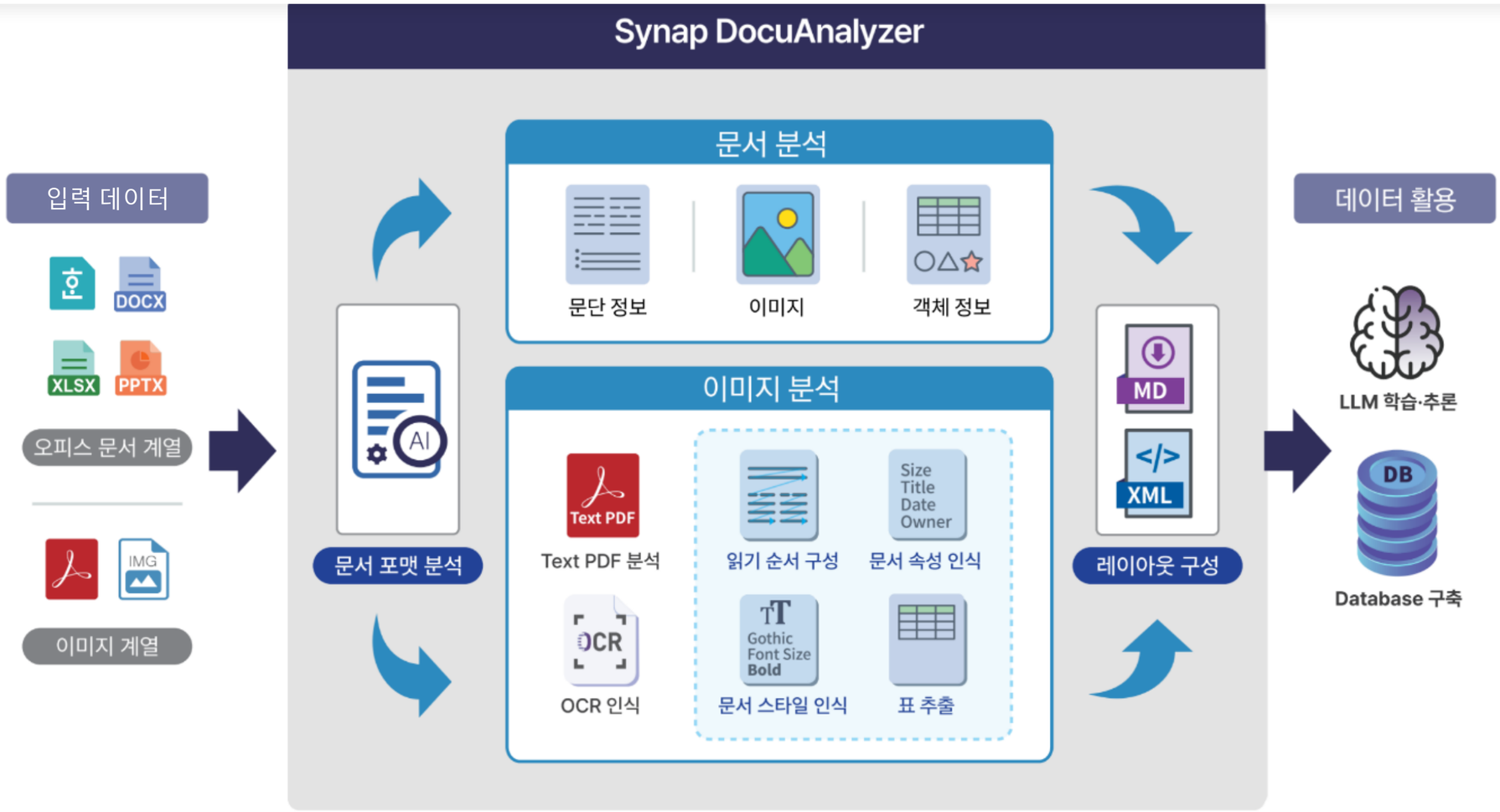
-다단 구조 문서 등 레이아웃이
복잡한 경우 인식 오류 발생

문서 구조와 시각적 정보를 인식하며, 비정형 문서를 정형 데이터로 변환할 수 있는 솔루션을 통해 해결!

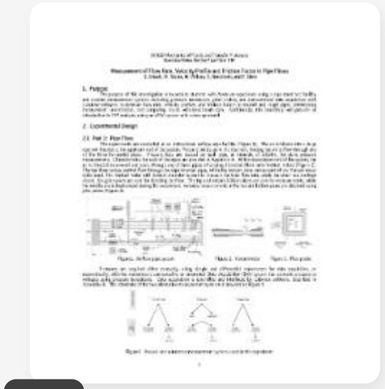
©2024. SYNAPSOFT. All rights reserved.

Synap DocuAnalyzer?

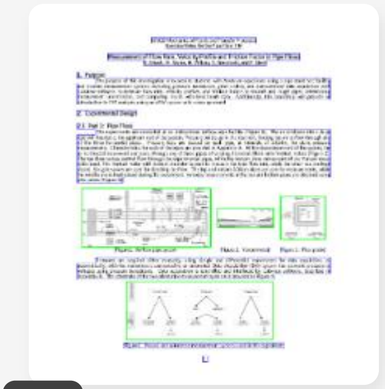
다양한 문서의 표, 이미지 등 시각적 정보와 복잡한 문서 구조 정보를 분석하여 Markdown 및 XML, JSON의 정형 데이터로 변환해주는 디지털 자산화 솔루션.
컴퓨터가 읽고 이해할 수 있는 정형 데이터는 RAG기술과 함께 LLM 구축에 필수적으로 요구됨



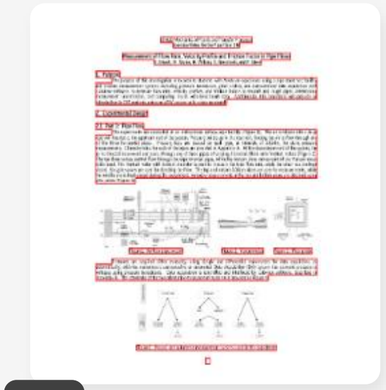
문서의 포맷, 레이아웃, 속성 및 객체 분석에 기반한 문서 구조 정보 인식



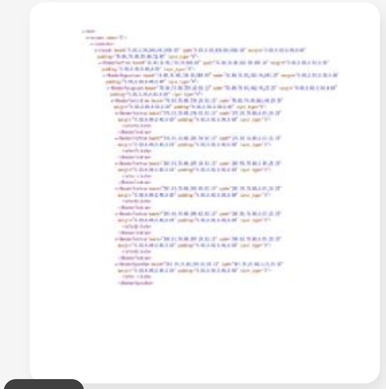
01 대상 문서



02 포맷, 레이아웃 분석

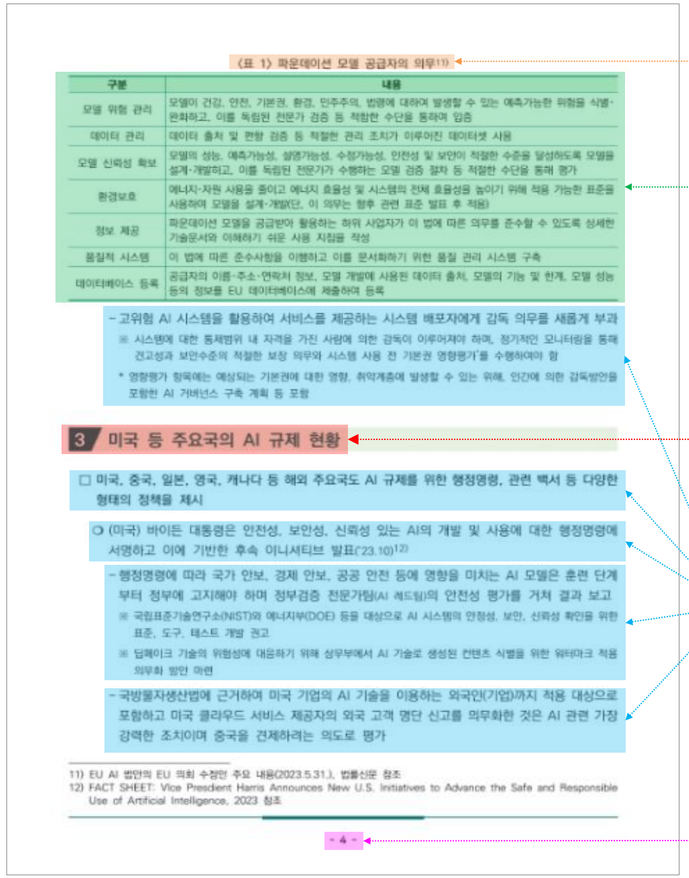


03 속성 및 객체 분석



04 문서 구조 추출 결과

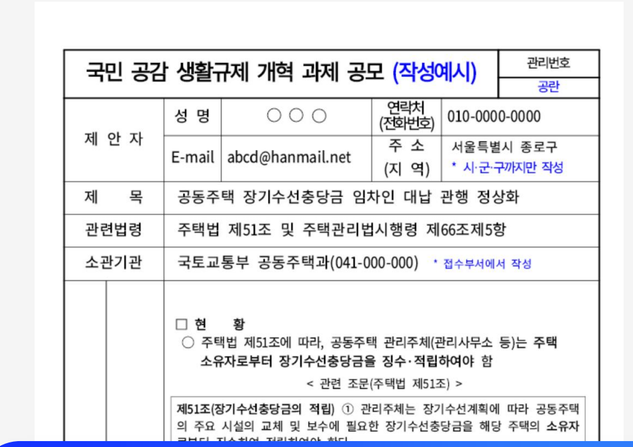
05 문서 구조 정보



©2024. SYNAPSOFT. All rights reserved.

synapsoft

비정형 문서를 XML, Markdown, JSON 포맷의 정형 데이터로 변환





문서 구조 정보



Input

비정형 문서

PDF부터 한글, 워드, 파워포인트,
이미지 등 다양한 포맷에
문단과 이미지, 표가 섞인 상태

Output

Markdown, XML, JSON 등
컴퓨터가 문서를 처리할 수 있는
포맷으로 변환

Advanced

DB 구축 및 AI 분석에 활용

기업의 문서 기반 Database 구축과
데이터에 기반한 RAG 및
LLM 구축에 활용

©2024. SYNAPSOFT. All rights reserved.

| 기업 내부 문서의 대부분은 일반 문서

Only Read

읽을 수만 있는 문서

PDF, 이미지 문서

<

Can Edit

읽고 수정, 변환 가능한 일반 문서

아래아한글(hwp,hwpx), 워드(doc,docx)
파워포인트(ppt,pptx),엑셀(xls,xlsx)



PDF 변환 없이 원본 문서 그대로를 분석하여 더 빠르고 손실 없는 100% 데이터 변환 수행

디지털 문서에 대해 수십년간 축적된 사이냅소프트의 노하우로 더 빠르고 정확한 분석이 가능합니다

©2024. SYNAPSOFT. All rights reserved.

일반 문서, 손실없이 구조 분석 가능

타사

일반문서
아래아한글, 워드, 파워포인트

→

PDF화
문서의 논리적 / 물리적 속성
영구 손실

→

OCR 및 구조 분석
좌표, 텍스트, 픽셀 단위로 분해된
물리적 속성을 재구성
더 많은 비용, 최종 품질 저하

→

결과물 출력
HTML, Markdown 등

사이냅 도큐애널라이저

일반 문서
아래아한글(hwp,hwpx), 워드(doc,docx)
파워포인트(ppt,pptx),엑셀(xls,xlsx)

→

구조 분석
✅ 문서의 논리적 / 물리적 속성을 그대로 활용
✅ 문단, 차트 등 최대한 반영, 재구성

→

결과물 출력
Markdown, XML, JSON 등
추출된 이미지는 별도 파일로 저장

문서 이해에 대한 근본 기술력으로 구조 분석 품질을 향상

©2024. SYNAPSOFT. All rights reserved.

PDF, 이미지 문서도 막힘 없는 구조 분석 및 결과물 출력

사이냅 도큐애널라이저

PDF, 이미지 문서
PDF, JPG, PNG, TIFF, BMP, GIF, PDF 등

→

OCR 및 구조 분석
✅ **Readerble PDF 정보와 OCR정보의 교차검증**
✅ 문단, 차트 등 최대한 반영, 재구성

→

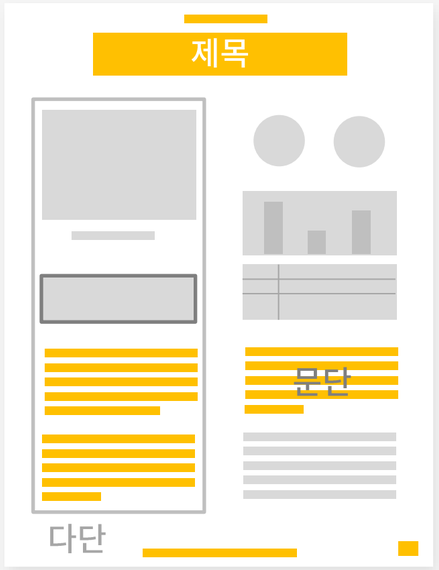
결과물 출력
Markdown, XML, JSON 등
추출된 이미지는 별도 파일로 저장

유통/보존을 위한 문서형식 (PDF, 이미지)도 꼼꼼히 분석하여 정형 데이터로 변환

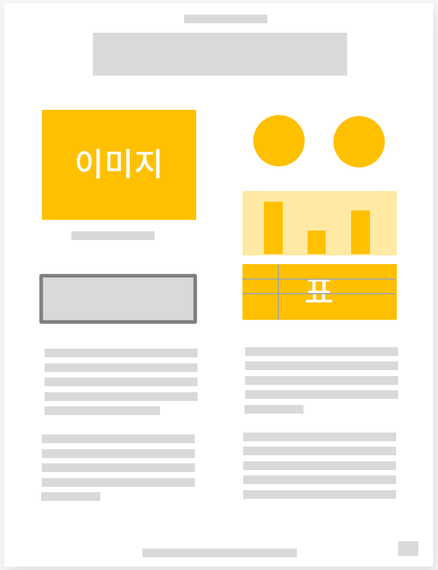
©2024. SYNAPSOFT. All rights reserved.

문서 구조, 시각 요소를 포함한 주요 분석 결과

LLM 모델 구축 및 RAG에 필요한 고품질 학습 데이터셋 구축 가능



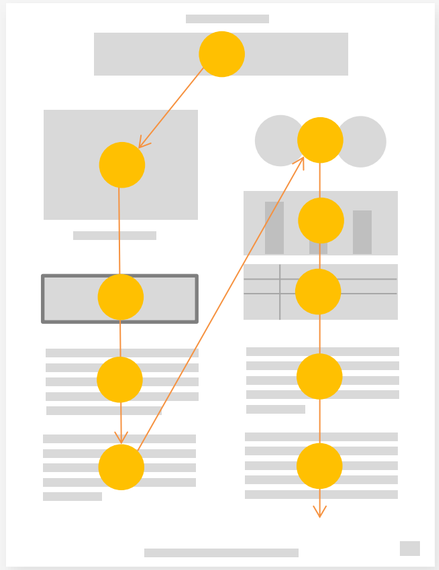
제목, 문단, 다단,페이지



표, 이미지, 차트 등
시각적 요소



이미지, 차트의 캡션(설명)



읽기 순서
(좌표 / 단 순서)

©2024. SYNAPSOFT. All rights reserved.

문서 포맷별 기능 지원





©2024. SYNAPSOFT. All rights reserved.

| 다양한 문서 지원과 정확한 문서 구조 분석으로 데이터 활용성 극대화

01

다양한 업무 문서 포맷 지원

한글 / MS Office / PDF / 이미지
폭넓은 문서 포맷 지원으로
기업 내 대부분의 주요 문서 포맷 대응



02

숨겨진 문서 구조 완벽 분석

제목, 문단, 머리글, 바닥글, 페이지
번호, 캡션, 리스트 등 상세 문서
구조정보의 인식 및 추출

표, 차트, 이미지 등
시각적 정보 인식 및 추출



03

활용도 높은 출력 형식

Markdown(마크다운) 출력 지원으로
대규모 언어모델(LLM) 구축 활용

XML, JSON 출력 지원으로
기업 Database 구축 활용



©2024. SYNAPSOFT. All rights reserved.

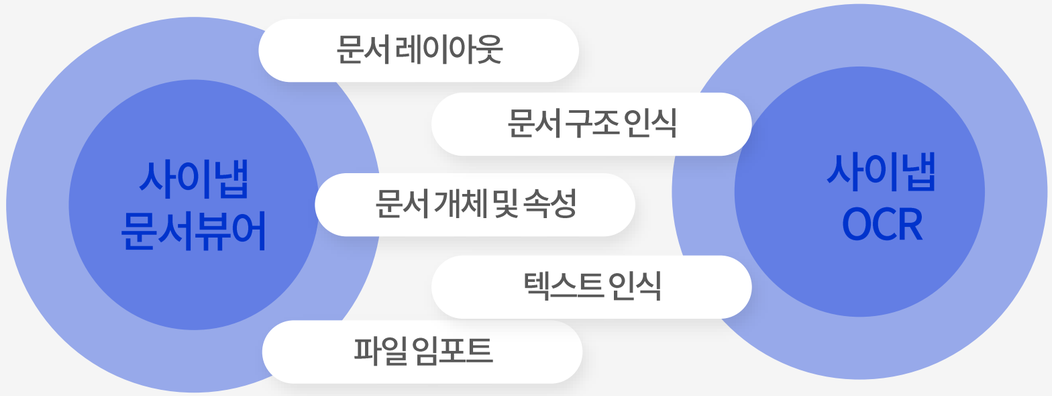
| 시장에서 검증된 사이냅 뷰어, 사이냅 OCR의 기술과 노하우가 반영된 도큐애널라이저

기업 고객 7000+

연동 솔루션 104+

파트너 288+

중앙행정부 85%+ 사용



대한민국 기관 및 기업에서 가장 많이 사용중인 문서뷰어,
한국정보통신기술협회 (TTA)인증 공공행정문서 인식률 99.3%의 OCR 기술이 적용

기업/기관 100+
국내 최다 레퍼런스



사이냅 도큐애널라이저의 신뢰성 보장

©2024. SYNAPSOFT. All rights reserved.

Synap DocuAnalyzer Spec

하드웨어

CPU
Intel Xeon 3GHz 8cores 16Threads 이상

Memory
64GB 이상

Storage(HDD)
최소 설치 공간 10GB 이상

GPU
24GB 이상의 메모리와 CUDA 코어가 탑재 NVIDIA GPU
(*이미지 문서 분석 시 필요)

지원 OS

x86 계열의 64bit 리눅스 시스템

- Debian-based : Ubuntu 20.04 x86_64
- Red Hat-based : CentOS 8 x86_64

제품 구성

설치 모듈, 라이선스, 매뉴얼



©2024. SYNAPSOFT. All rights reserved.

synapsoft

| 복잡한 문서 구조 분석 단순하게! 데이터의 신뢰성은 높이고, 비용 절감 효과까지!

01

생성형 AI 신뢰성 강화

근거 문서 기반 답변, RAG 모델 정확도 향상
생성형 AI의 허위 정보 생성 (할루시네이션) 현상 최소화에 활용
다양한 형태의 문서 데이터 정형화 > RAG 모델의 데이터 활용 범위 확대

02

문서 분석 프로세스 단순화

다양한 포맷의 분석, 분류 과정 한 번에 처리
텍스트 추출, 문단 구분, 표 인식 등
복잡한 작업을 솔루션 하나로 해결

03

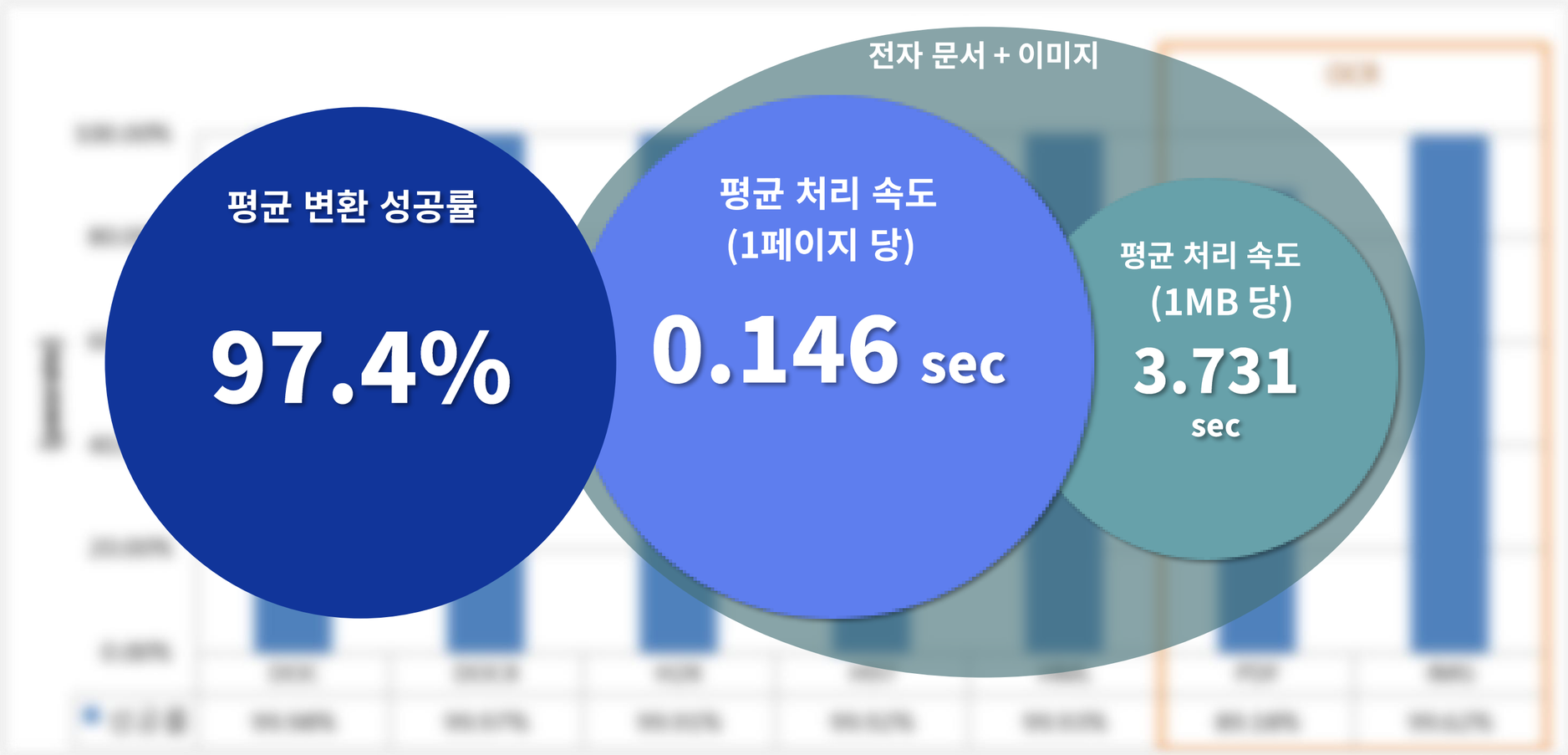
리소스 절감 및 효율성 향상

향상된 정보 접근성 기반 업무 자동화
대화형 검색으로 정보 접근 효율성 대폭 향상
LLM 구축 시 데이터 전처리 자동화로
시간 / 비용 절감

©2024. SYNAPSOFT. All rights reserved.

안정적인 속도와 정확성

내부 테스트 결과 : 72,860건의 일반 문서를 수집하여 모듈 신뢰성 테스트 진행



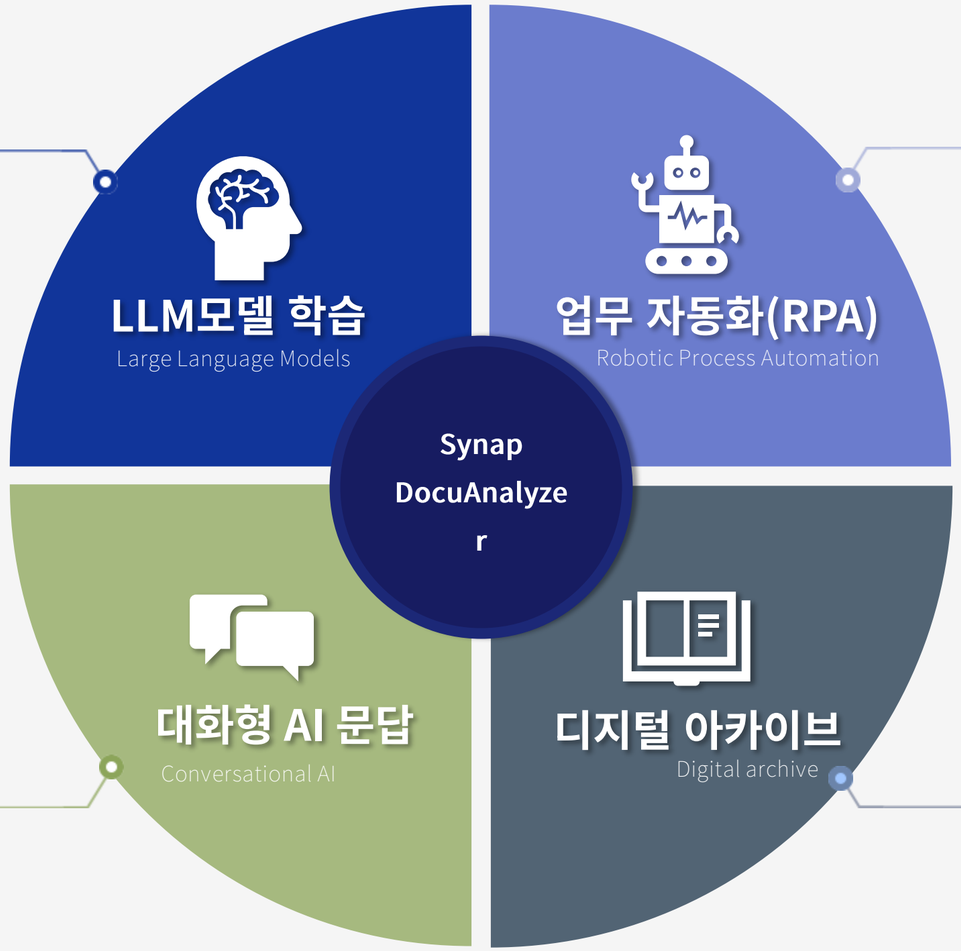
테스트 장비 환경 :
- Module : Ubuntu 22.04.2 LTS, i5-4690@3.50GHz, Memory 16GB
- OCR Engine : Ubuntu18.04, i9-10900K@3.70GHz, RTX 3090(24GB), Memory : 125GB

©2024. SYNAPSOFT. All rights reserved.

혁신적인 문서 구조 분석 기술로 다양한 분야에 활용 가능

보유 문서 기반 최적화된 LLM 모델 구축

- 다양한 형태의 문서 이해
- 도메인 특화 지식 통합
- 고품질 데이터 셋 구축
- LLM 모델 성능 최적화



향상된 정보 접근성 기반 업무 자동화 실현

- 필요 데이터 추출 및 입력 자동화
- 데이터베이스 구축으로 업무 자동화 시스템 구현
- 데이터 접근 권한 관리 기술에 활용 가능

고도화된 대화형 AI 개발

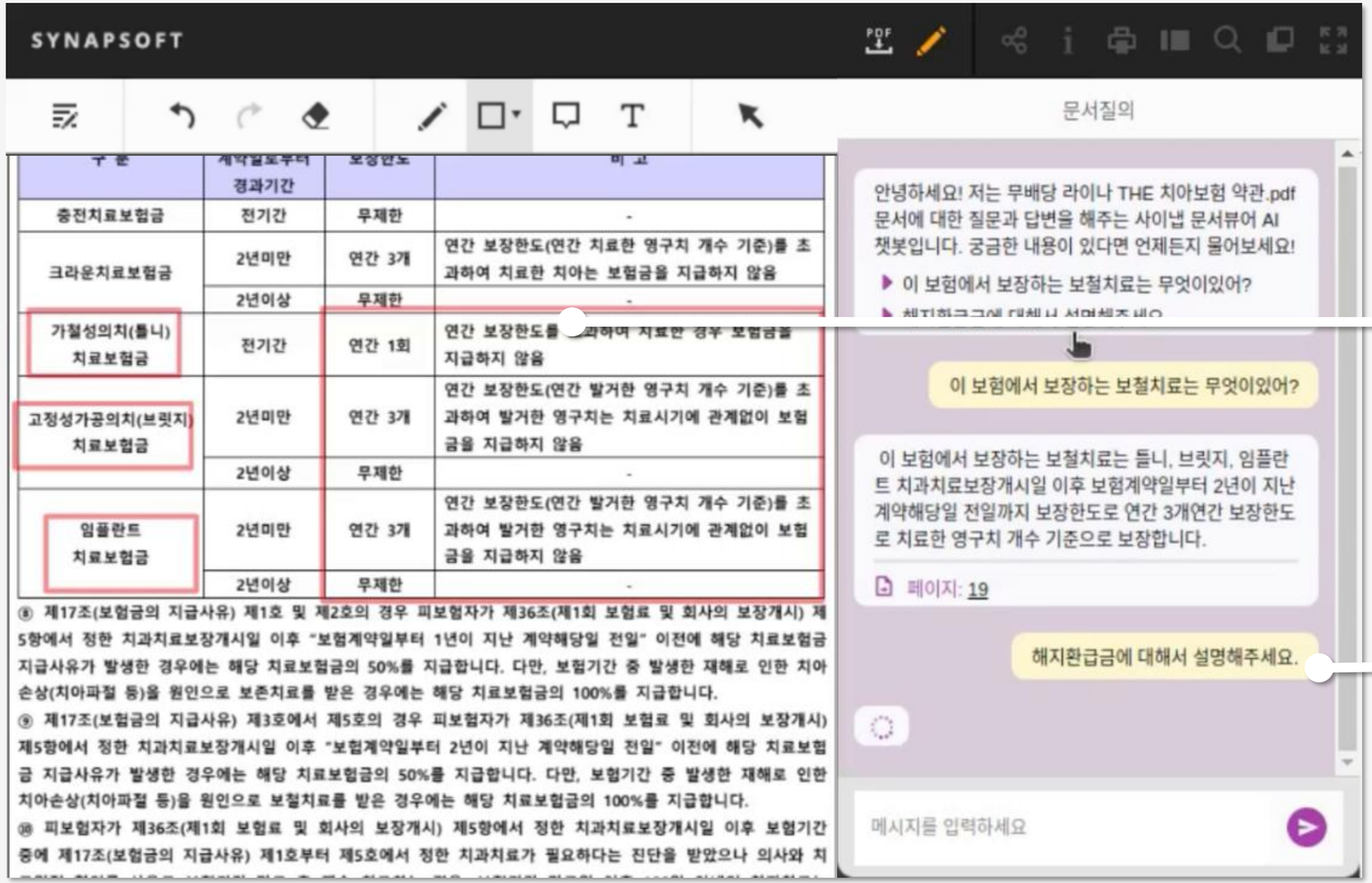
- 표, 이미지 등 다양한 콘텐츠 이해
- AI 자연어 처리(NLP) 능력 향상
- 풍부한 정보 기반 복잡한 사용자 질문 대응

최적화 대규모 문서 디지털 아카이브 구축

- 메타 데이터 생성•관리
- 비정형 문서의 정형화
- 자동 문서 분류 및 색인화 기술에 활용 가능

©2024. SYNAPSOFT. All rights reserved.

| 사이냅 문서 뷰어 > 문서 기반 질문 답변 기능에 활용



구분	계약일로부터 경과기간	보장한도	비고
충전치료보험금	전기간	무제한	-
크라운치료보험금	2년미만	연간 3개	연간 보장한도(연간 치료한 영구치 개수 기준)를 초과하여 치료한 치아는 보험금을 지급하지 않음
	2년이상	무제한	-
가철성의치(틀니) 치료보험금	전기간	연간 1회	연간 보장한도를 초과하여 치료한 경우 보험금을 지급하지 않음
고정성가공의치(브릿지) 치료보험금	2년미만	연간 3개	연간 보장한도(연간 발거한 영구치 개수 기준)를 초과하여 발거한 영구치는 치료시기에 관계없이 보험금을 지급하지 않음
	2년이상	무제한	-
임플란트 치료보험금	2년미만	연간 3개	연간 보장한도(연간 발거한 영구치 개수 기준)를 초과하여 발거한 영구치는 치료시기에 관계없이 보험금을 지급하지 않음
	2년이상	무제한	-

⑧ 제17조(보험금의 지급사유) 제1호 및 제2호의 경우 피보험자가 제36조(제1회 보험료 및 회사의 보장개시) 제5항에서 정한 치과치료보장개시일 이후 "보험계약일부터 1년이 지난 계약해당일 전일" 이전에 해당 치료보험금 지급사유가 발생한 경우에는 해당 치료보험금의 50%를 지급합니다. 다만, 보험기간 중 발생한 재해로 인한 치아손상(치아파절 등)을 원인으로 보존치료를 받은 경우에는 해당 치료보험금의 100%를 지급합니다.

⑨ 제17조(보험금의 지급사유) 제3호에서 제5호의 경우 피보험자가 제36조(제1회 보험료 및 회사의 보장개시) 제5항에서 정한 치과치료보장개시일 이후 "보험계약일부터 2년이 지난 계약해당일 전일" 이전에 해당 치료보험금 지급사유가 발생한 경우에는 해당 치료보험금의 50%를 지급합니다. 다만, 보험기간 중 발생한 재해로 인한 치아손상(치아파절 등)을 원인으로 보철치료를 받은 경우에는 해당 치료보험금의 100%를 지급합니다.

⑩ 피보험자가 제36조(제1회 보험료 및 회사의 보장개시) 제5항에서 정한 치과치료보장개시일 이후 보험기간 중에 제17조(보험금의 지급사유) 제1호부터 제5호에서 정한 치과치료가 필요하다는 진단을 받았으나 의사와 치

문서 구조를 LLM에 활용할 수 있는 데이터 형태로 변환 적용

질의를 통해 복잡한 보험 약관에서도 빠르게 원하는 정보 검색

©2024. SYNAPSOFT. All rights reserved.

사이냅소프트



인공지능 디지털 문서 SaaS 전문 기업

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

©2024. SYNAPSOFT. All rights reserved.

Synap **DocuAnalyser**

데이터의 신뢰성과 정확성을 극대화하는 문서 구조 분석 전문 솔루션

제품 문의

사이냅소프트 솔루션사업팀

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

서울특별시 강서구 마곡중앙4로 62

www.synapsoft.co.kr

synapsoft

©2024. SYNAPSOFT. All rights reserved.