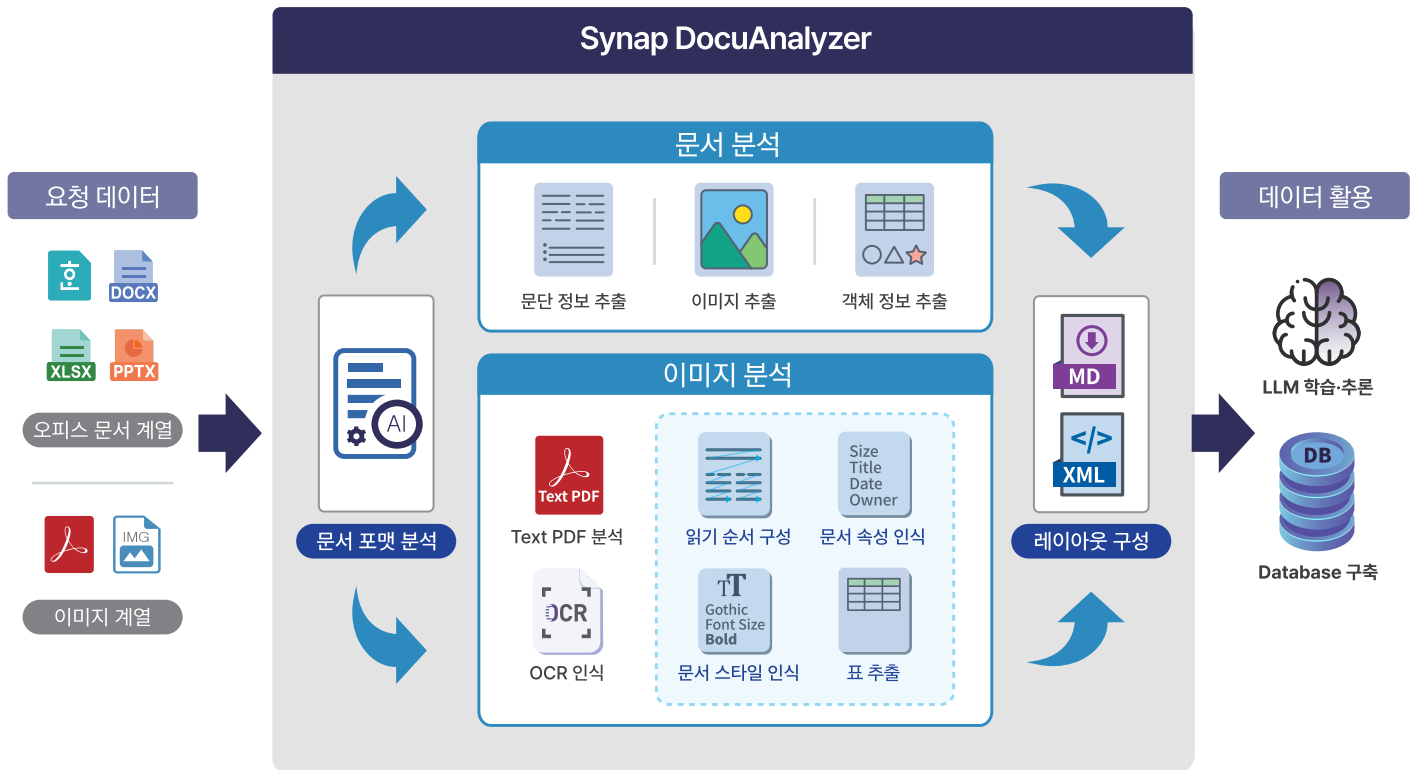


문서 구조 분석 전문 솔루션 사이냅 도큐애널리저

다양한 문서에서 표, 이미지 같은 시각적 정보와 복잡한 문서 구조 정보를 분석하여
Markdown 및 XML 형식의 정형 데이터로 변환해 주는 디지털 문서 자산화 필수 솔루션



특장점



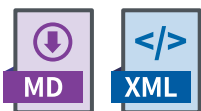
업무 문서 포맷에 대한 높은 접근성

- 한글 / MS-Office / PDF / 이미지까지 폭넓은 문서 포맷 지원으로 기업이 보유한 대부분의 문서 형식 대응 가능



숨겨진 문서 구조 완벽 분석

- 제목, 문단, 머릿글, 바닥글, 페이지 번호, 캡션 등 상세 문서 구조정보의 인식 및 추출
- 표, 이미지 등 시각적 정보 인식 및 추출



활용도 높은 출력 형식

- Markdown(마크다운) 출력 지원으로 대규모 언어 모델(LLM) 구축 활용
- XML 출력 지원으로 기업 Database 구축 활용

주요 기능

다양한 지원 포맷



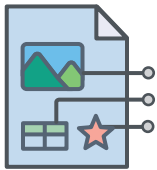
- 입력 포맷 : 아래아한글 / MS Office / PDF / 이미지
- 출력 포맷 : XML, Markdown

문서 구조 분석



- 제목, 문단
- 머리글, 바닥글
- 페이지 번호
- 캡션 (표/이미지 등의 제목)

시각적 정보 분석



- 표 (병합정보)
- 이미지
- 차트

읽기 순서 지원



- 문서 읽기 순서 지원
- 읽기 순서 기반 레이아웃 구성
- 다단 구조 인식 및 처리

적용 분야

LLM모델 학습

다양한 포맷의 문서 내용 이해 및 보유 문서 기반 LLM모델 구축

대화형 AI 문답

표, 이미지 등 문서 이해로 자연어 처리 능력 향상 및 대화형 AI 문답 시스템 개발

업무 자동화(RPA)

필요 정보만 추출하여 반복적 업무 프로세스 자동화 시스템 구축

디지털 아카이브

비정형 문서의 정형화 및 대규모 문서의 디지털 아카이브 구축

기대효과

문서 구조 분석 과정의 단순화! 정보 가치는 극대화하고 비용은 절감!

문서 분석 프로세스 단순화

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

생성형 AI의 신뢰성 강화

- 근거 문서 기반 답변으로 정확도 상승
- 생성형 AI의 허위정보 생성여부 판단에 활용

리소스 절감 및 효율성 향상

- 향상된 정보 접근성 기반 업무 자동화
- 대화형 검색으로 정보 접근 효율성 향상