

기관 서버 안에서, HWP 문서를 RAG용 데이터로.

Luminir Document Parser는 르몽이 만든 설치형(온프레미스) 문서 파서입니다. HWP·HWPX·PDF·워드·파워포인트·엑셀·스캔 이미지 문서에서 제목·본문·표·그림 영역을 찾아 읽기 순서대로 Markdown과 JSON으로 바꿉니다. 기관 서버에 설치해 RAG(문서를 검색해 답하는 AI) 시스템에 넣을 데이터를 만드는 데 씁니다.

도입 방식

온프레미스 (기관 서버에
설치)

설치 형태

컨테이너(Docker)

운영체제

리눅스

GPU

NVIDIA GPU

작동 방식

01

문서 받기

HWP·HWPX, PDF, 워드·파워포인트·엑셀 문서, 이미지 등 34종 파일을 그대로 받습니다. 형식마다 따로 바꿀 필요가 없습니다.

02

레이아웃 분석

쪽마다 제목, 본문, 표, 그림, 머리말 영역을 찾고 읽기 순서를 정합니다.

03

영역별 인식

본문은 글자로, 표는 병합 셀까지 행·열 구조로, 수식은 LaTeX(수식 표기 형식)로 읽습니다.

04

결과 내보내기

Markdown과, 요소마다 종류·쪽 번호·위치 좌표가 담긴 JSON으로 돌려줍니다. 머리말과 쪽 번호에는 종류가 표시돼 있어 필요하면 쉽게 뺄 수 있습니다.

넣는 문서

HWP·HWPX 한글 문서를 PDF로 바꾸지 않고 바로 처리

PDF 글자가 있는 PDF와 스캔 PDF 모두

워드·발표 DOC·DOCX·RTF·ODT, PPT·PPTX·ODP

엑셀·CSV XLS·XLSX·ODS·CSV·TSV를 시트 구조 그대로 표로

이미지 PNG·JPG·TIFF·BMP·WEBP (사진, 스캔)

텍스트 TXT·MD·HTML·XML·JSON 등

받는 결과

Markdown 제목 단계, 표, 그림 위치를 살린 텍스트

JSON 요소 종류, 쪽 번호, 위치 좌표, 읽기 순서

TXT 구조 없이 본문만 필요할 때

API 업로드부터 결과 조회까지 REST API로 연동

업무에 맞춘 처리

HWP·HWPX 직접 처리

제목 단계 출력

머리말·쪽 번호 구분

요소별 위치 좌표

여러 파일 일괄 변환과 중단 후 이어서 처리

폐쇄망 설치 구성

공개 벤치마크 (르몽 자체 측정)

벤치마크	데이터	점수	비교	측정일
opendataloader-bench	한컴 공개 · 영어 문서 200쪽	0.914	공개 1위 점수(0.907)보다 높음	2026-10-12
KoOCR-Bench	schift-io v1.2 · 한국어 공공문서 623쪽	0.759	공개 리더보드 상위권	2026-10-12
DP-Bench	Upstage 공개 · 영어 문서 200쪽	0.962	함께 비교한 상용 파서 5종보다 높음	2026-09-23
OmniDocBench v1.5	문서 1,651쪽	0.892	—	2026-09-19
FinTabNet	재무제표 표 300개	0.948	—	2026-10-12

모든 점수는 각 벤치마크가 공개한 데이터와 채점 코드로 르몽이 직접 측정했습니다(자체 측정). 결과 원본은 요청하시면 드립니다. 측정 조건 상세: <https://luminir-landing.lemong.ai/document-parser/benchmarks/>

보안·설치

온프레미스·폐쇄망 설치 NVIDIA GPU가 있는 리눅스 서버에 컨테이너로 설치합니다.

외부 AI 연결 없이 운영 레이아웃 분석과 인식 모델을 기관 서버에 함께 설치해, 인터넷 연결 없이 동작하도록 구성합니다.

권한과 기록 관리자·사용자 권한 구분, 문서 암호화 저장, 처리 이력 감사 기록을 지원합니다.

활용 분야

규정·지침 질의응답 규정집, 업무 지침, 매뉴얼을 조항과 표 구조를 살린 채 검색 시스템에 넣어, 질문하면 해당 조항을 찾아 답하게 합니다.

쌓여 있는 HWP 문서 정리 수년간 쌓인 한글 문서를 PDF로 바꾸지 않고 한꺼번에 변환합니다. 중간에 멈춰도 멈춘 곳부터 이어서 처리합니다.

통계표·보고서 데이터화 병합 셀이 많은 통계표를 행·열 구조 그대로 꺼내 엑셀이나 데이터베이스로 옮깁니다.

원문 위치 표시 요소마다 쪽 번호와 위치 좌표가 있어, 검색 결과나 AI 답변에 쓰인 문단을 원본 문서 위에 표시할 수 있습니다.

도입 절차

도입 상담 → 기관 서버에 시범 설치 → 실제 문서로 결과 확인 → 계약과 설치·연동 순서로 진행합니다. 상담에서 문서 종류와 양, 서버 환경을 확인하고, 기간과 조건을 안내해 드립니다.

르몽 주식회사

서울 서초구 양재대로12길 56, AI캠퍼스 A동 603호

회사 실적: 2025년 9월 한국전력 AI 기반 사내시스템 구축 용역 수주

도입 상담

support@lemong.ai

02-6958-6868

<https://luminir-landing.lemong.ai/document-parser/>