

한국 공공기관·정부 대상 제안서

CJK 문자처리 및 공공 AI 전처리 API/SDK 「CAW」

공공문서, 기록물, 민원, 교육·연구 데이터에 포함된 한자·일본어·중국어·한국어 혼합 문자를 정확히 입력·분석·검색
·AI 전처리하기 위한 공공업무용 기술 제안

CAW는 CJK 입력 지원, 한자 구조 분석, 검색 전처리, 공공 AI/RAG 전처리를 API/SDK 형태로 제공하여 공공기관의 문서·기록물·민원·교육 데이터 활용도를 높이는 B2G/B2B 기술 모듈입니다.

적용 대상 기관

- 중앙부처·지방자치단체: 민원, 행정문서, 정책자료, 기록물 검색·분류 고도화
- 공공기관·공기업: 업무시스템 내 다국어·한자 포함 문서 처리 효율화
- 교육청·대학·연구기관: 한자·일본어·중국어 교육 및 연구데이터 처리
- 도서관·기록원·문화기관: 고문서, 역사자료, 한자 포함 콘텐츠 검색 지원
- 공공 SI·시스템 운영사: 기존 공공시스템에 CJK 문자처리 모듈 연동

공공업무 가치	기대 효과
CJK 문자처리 기능을 기관별 시스템에 API/SDK로 추가	신규 개발기간 단축 및 도입 리스크 완화
한자·일본어·중국어 포함 문서의 입력·검색·분류 지원	행정문서·기록물·민원 검색 품질 향상
공공 AI/RAG 구축 전 데이터 전처리 계층으로 활용	AI 답변 품질 및 공공데이터 활용성 향상
온프레미스·전용 API·SDK 납품 등 기관 환경에 맞춤 제공	망분리·보안정책·운영환경에 대한 대응 가능성 확보

※ 본 제안서는 기존 해외 SW 기업 대상 CAW API/SDK 제안 내용을 한국 공공기관·정부 도입 검토 형식으로 재구성한 문서입니다.

1. 공공기관의 문제 인식과 CAW 의 해결 방향

공공기관의 문서와 데이터에는 한글, 한자, 일본어, 중국어, 영문, 숫자, 고유명사, 지명, 기관명, 전문용어가 혼합되어 존재합니다. 특히 기록물, 고문서, 민원, 교육·연구 데이터, OCR 결과물, 공공 AI 학습·검색 데이터에서는 CJK 문자 특성이 검색 품질과 업무 효율에 직접적인 영향을 줍니다.

공공업무 현장의 문제	CAW 적용 방향
한자·일본어·중국어가 포함된 기록물과 행정문서 검색이 어렵고, 표기 변형으로 누락이 발생	한자 구조 분석과 CJK 전처리 API 로 검색어 확장·정규화·분류 보조
OCR 또는 수기 입력 데이터에 한자·이체자·전문용어가 섞여 문서 정제가 오래 걸림	문자 구조 기반 전처리로 정제·검수·분류 업무를 보조
공공 AI/RAG 구축 시 한국어·한자·외국어 혼합 데이터의 전처리 품질이 일정하지 않음	AI 전처리 계층으로 문서 분절, 용어 처리, 검색 입력 품질 개선 지원
기관별 업무시스템에 CJK 기능을 직접 개발하기에는 예산·기간·전문성이 부담	API/SDK 또는 설치형 모듈로 기존 시스템에 단계적 연동

2. CAW API/SDK 개요

CAW API/SDK 는 한국어·한자·일본어·중국어 등 CJK 문자 환경에서 입력, 분석, 검색, AI 전처리를 지원하는 공공업무용 문자처리 엔진입니다. 단순 입력기 또는 번역기가 아니라, 공공문서와 데이터에 포함된 CJK 문자를 시스템과 AI 가 활용하기 쉬운 형태로 분석·정규화·연동하는 보조 기술입니다.

기능	설명
CJK 입력 지원	한자·일본어·중국어·한국어 혼합 입력 및 변환 흐름 보조
한자 구조 분석	부수, 구성요소, 조합 패턴 등 한자 구조 정보를 분석
검색 전처리	표기 차이, 한자 포함 검색어, 문서 색인 전처리 보조
공공 AI 전처리	RAG, LLM, AI 검색에 활용할 CJK 데이터 정제·분석 계층 제공
SDK 연동	웹, 모바일, 업무시스템, 공공 SI 환경에 모듈 형태로 연동
API 서비스	기관별 API Key, 사용량 관리, 로그 관리, 전용 환경 구성 가능

3. 공공기관 적용 시나리오

기관/업무 유형	적용 예시
기록물·문서관리 기관	한자·일본어·중국어가 포함된 기록물, 고문서, 스캔 문서의 검색·분류 보조
민원·행정서비스 부서	민원 문서, 질의응답, 행정용어 검색에서 CJK 표기와 고유명사 처리 지원
공공 AI/RAG 구축 부서	업무 매뉴얼, 정책자료, 법령·지침, 기술문서의 AI 검색 전처리
교육청·대학·연구기관	한자·일본어·중국어 학습 콘텐츠, 연구자료, 동아시아 문헌 데이터 처리
문화·관광·도서관 기관	문화유산, 관광안내, 도서·자료 검색에서 다국어·한자 표기 보조
공공 SI·운영사	기존 공공업무시스템에 CJK 문자처리 API/SDK를 내장하여 기능 고도화

4. 제공 방식

제공 방식	내용
기관 전용 API	기관별 API Key, 접근권한, 호출량 관리, 로그 모니터링 제공
SDK 납품	웹·모바일·업무시스템에 내장 가능한 SDK 또는 모듈 제공
온프레미스/폐쇄망	기관 보안정책에 따라 내부망 또는 전용 서버 구축 방식 검토
클라우드/SaaS	민간클라우드 또는 전용 클라우드 환경 구성 가능. 공공기관 요건은 별도 협의
SI 연계	공공 SI 사업자와 함께 기존 시스템에 CAW 모듈을 연동하는 방식

5. 공공기관 PoC 제안

초기 도입 단계에서는 4~8 주 범위의 공공업무 적용성 검증(PoC)을 권장합니다. PoC는 기관의 실제 데이터 전체를 요구하지 않고, 비식별화된 샘플 데이터 또는 공개 가능한 업무 시나리오를 기준으로 진행할 수 있습니다.

- 검증 대상: 기록물 검색, 민원/행정문서 검색, 한자 구조 분석, 공공 AI/RAG 전처리 등
- 검증 방식: 샘플 데이터 분석, API/SDK 연동 데모, 정량·정성 평가
- 결과물: 기술검증 보고서, 연동 데모, 도입 가능성 평가, 본사업 견적·로드맵
- 도입 목표: 파일럿 검증 후 기관 전용 API, SDK 납품, SI 연계 또는 설치형 구축으로 전환

6. PoC 범위·일정·산출물

기간	주요 활동
Week 1	업무 요구사항 확인, 대상 데이터/시나리오 선정, 보안·비식별 처리 방식 협의
Week 2-3	CAW API/SDK 기능 매핑, 샘플 데이터 전처리, 초기 연동 데모 구성
Week 4	1 차 검증: 검색·분석·AI 전처리 적용 결과 확인 및 개선사항 정리
Week 5-6	추가 기능 검증, 평가 지표 정리, 기관 업무 적용성 검토
Week 7-8	최종 보고, 본사업 전환 범위·예산·일정·운영 구조 제안

주요 산출물

산출물	내용
PoC 범위정의서	검증 대상 업무, 데이터 범위, 평가 기준, 보안 유의사항 정리
API/SDK 연동 데모	기관 업무 시나리오에 맞춘 화면 또는 API 호출 예시
기술검증 보고서	적용 결과, 개선 가능성, 제약사항, 상용도입 조건 정리
운영·보안 검토자료	데이터 처리, 접근통제, 로그, 종료·삭제 절차 설명
본사업 전환 제안서	도입 범위, 일정, 예산 산정, 유지보수 방식 제안

7. 예산 산정 예시

아래 금액은 공공기관 검토용 예산 산정 예시이며, 실제 견적은 데이터 범위, 연동 시스템 수, 설치 방식, 보안 요구사항, 유지보수 범위에 따라 조정됩니다.

구분	범위	예산 예시
사전 기술진단	1~2 주, 데이터/업무 적합성 검토, 적용 가능성 보고	100 만~300 만원
공공 Lite PoC	3~4 주, 기능 1~2 개 검증, 간단 데모 및 보고서	500 만~1,500 만원
공공 Standard PoC	4~8 주, API/SDK 연동 데모, 기술검증 보고서, 본사업 견적	1,500 만~3,000 만원
기관 전용 API/SDK 구축	전용 API 환경, SDK 연동, 관리자 기능, 기술 지원	3,000 만원~개별견적
온프레미스/폐쇄망 구축	기관 내부망 또는 전용 서버 설치, 보안요건 반영	개별견적
연간 유지보수	기술지원, 오류수정, 업데이트, 운영 리포트	계약금액의 10~15% 수준

8. 계약·보안·종료 처리

공공기관 도입에서는 기능 검증만큼 데이터 처리, 접근통제, 종료 시 조치, 원천기술 보호, 유지보수 범위를 명확히 하는 것이 중요합니다. CAW는 기관별 보안정책과 구매절차에 맞춰 다음과 같은 운영 원칙을 적용할 수 있습니다.

항목	처리 원칙
데이터 처리	PoC 단계에서는 비식별 샘플 데이터 또는 공개 가능한 데이터 사용을 우선하며, 원본 데이터 반출은 최소화
접근통제	기관별 API Key, 계정권한, IP 제한, 로그 기록 등 접근관리 방식 적용
소스코드/원천기술	CAW 핵심 엔진 소스코드는 제공하지 않으며, API/SDK 사용권·연동 모듈·문서 중심으로 제공
계약 종료	API Key 비활성화, 계정 정리, SDK 사용권 종료, 미사용 데이터 반환·삭제, 기술자료 폐기 절차 수행
운영 로그	필요 시 기관별 사용량, 오류, 처리 결과 로그를 제공하고 보관 기간은 계약서에 명시
전환 기간	업무 중단을 줄이기 위해 필요 시 30~90 일의 전환 기간을 별도 협의

9. 도입 경로 및 다음 단계

- 1 단계: 30 분 기술 설명 미팅 및 기관 업무 시나리오 확인
- 2 단계: 비식별 샘플 데이터 또는 공개 데이터 기준 사전 기술진단
- 3 단계: 4~8 주 공공 Standard PoC 수행
- 4 단계: 본사업 전환 범위·예산·조달·SI 연계 방식 검토
- 5 단계: 기관 전용 API/SDK 또는 설치형 구축 및 유지보수 계약 추진

공공기관 제안 문구

CAW는 공공기관의 한자·일본어·중국어·한국어 혼합 문서와 데이터를 대상으로 입력, 검색, 분류, AI 전처리를 지원하는 API/SDK입니다. 초기에는 비식별 샘플 데이터 기반 PoC로 적용성을 검증하고, 이후 기관 전용 API·SDK·설치형 구축으로 확장할 수 있습니다.

문의: [회사명] / [담당자명] / [이메일] / [URL] / Service: CAW API/SDK

※ 본 문서는 공공기관 검토용 제안 초안입니다. 실제 제안·견적·계약 조건은 기관의 과업범위, 보안요건, 구매절차, 연동 대상 시스템에 따라 별도 협의가 필요합니다.