RAG chunking: 문서를 얼마나 잘게 나눠야 하는가?

chunking의 정의와 필요성, 작동 원리, SAG 아키텍처 적용 기준과 실무 체크리스트를 연구·공식 문서 근거로 설명합니다.

어떤 상황에서 활용하나요?

문서 검색·근거 선택의 정확성과 비용을 비교하는 AI 개발자가 설계 대안을 검토할 때 읽는 연구 설명입니다. 평가 질문·검색 후보·출처와 실패 사례를 기준으로 적용 여부를 결정하며, 소개한 연구 알고리즘이 SAG에 모두 구현됐다는 의미는 아닙니다.

SAG(새그)는 더크림유니언이 운영하는 GEO·AEO·SEO 분석 및 최적화 플랫폼입니다. 검색엔진과 생성형 AI가 브랜드와 웹사이트 콘텐츠를 어떻게 발견하고 이해하는지 분석하고, 경쟁사 비교·개선 우선순위·실행 결과·재검증을 관리합니다. 플랫폼 구독과 전문 운영 지원을 제공하며, 콘텐츠 제작·사이트 코드 수정 등 대행 실행 범위는 별도 계약에서 정합니다.

한 문장 정의

chunking은 긴 문서를 검색 가능한 의미 단위로 나누고 원문 위치와 계층을 보존하는 과정입니다.

> 핵심 답: 너무 큰 조각은 관련 없는 문맥을 섞고 너무 작은 조각은 조건과 예외를 잃습니다. 글자 수만으로 자르면 표, 제목, 문단 관계도 끊어집니다.


공식 URL과 페이지 근거를 대조하는 SAG 분석 예시

> 이 화면은 방법 설명을 위한 SAG 공개 예시이며, 특정 고객사의 성과나 색인·검색·AI 추천을 보장하지 않습니다.

왜 이 기술이 필요한가?

너무 큰 조각은 관련 없는 문맥을 섞고 너무 작은 조각은 조건과 예외를 잃습니다. 글자 수만으로 자르면 표, 제목, 문단 관계도 끊어집니다.

작동 원리

제목 계층과 문단, 표, 목록을 기준으로 먼저 나눈 뒤 길이 제한과 overlap을 적용합니다. 각 조각에는 문서·섹션·버전·원문 위치를 메타데이터로 둡니다.

설계할 때는 정확도만 보지 않습니다. 지연 시간, 비용, 데이터 경계, 갱신 주기와 실패 시 동작을 함께 정의해야 운영에서 재현 가능한 결과가 됩니다. 자동화가 확신하지 못하는 값은 0이나 성공으로 바꾸지 않고 미측정·검토 필요 상태로 남기는 것이 안전합니다.

SAG 기술과의 연결

SAG의 근거 artifact에 chunk를 연결한다면 보고서에서 원문 페이지와 구간으로 돌아갈 수 있어야 합니다. 현재 저장하는 출처·해시·버전 원칙이 그 기반입니다.

실무 체크리스트

- 문서 구조를 우선해 경계를 정합니다
- 문맥 손실과 중복 검색률을 함께 측정합니다
- chunk에서 원문으로 돌아가는 링크를 보존합니다
- 실패·빈 결과·권한 오류의 상태를 성공과 구분합니다
- 변경 전후를 동일한 조건으로 재검증합니다

연구와 공식 문서

- RAG 원 논문

참고 문서는 원리와 권장사항의 근거입니다. 검색 노출, AI 언급, 순위나 매출을 보장하지 않으며 실제 적용 효과는 서비스 데이터와 동일 조건의 관측으로 확인해야 합니다.

선택 기준과 구체적인 적용 예

사양표를 숫자 행만 남겨 잘라내면 단위·제품명·시험 조건이 사라집니다. 제목과 표의 머리글을 함께 보존하고 원문 위치를 연결하세요. 모든 문서에 적용되는 최적 조각 크기는 없으며 문서 유형과 평가 질문으로 선택합니다.

SAG에서의 적용 범위

이 글은 검색 AI의 연구 원리와 확장 설계를 다룹니다. SAG의 페이지 수집·근거 기록·보고서 검증 구조와 연결해 읽되, 논문의 검색 알고리즘이 모두 운영 파이프라인에 탑재됐다는 의미로 해석하지 않습니다. 적용 여부는 검색 모듈, 평가 데이터와 실행 기록으로 확인합니다.

이 기술을 이어서 읽는 방법

RAG·GraphRAG·Self-RAG 논문과 적용 조건을 비교합니다.

- 검색 AI의 확장 원리
- 기능 안내 FAQ