robots.txt와 접근 권한: 수집 가능한 것과 수집해도 되는 것은 같은가?

자동 접근 지시와 소유자의 허용 범위를 함께 적용하는 운영 규칙입니다. 기술적으로 페이지를 열 수 있어도 수집 권한과 사용 범위가 일치한다고 단정할 수 없습니다. 제한을 실패 점수로 바꾸면 분석도 왜곡됩니다.

어떤 상황에서 활용하나요?

공식 사이트 자료를 등록하거나 수집 실패·반복 수집을 검토하는 웹 운영 담당자에게 유용합니다. 보유한 자료와 접근 권한을 먼저 확인하고, 실제 읽은 본문과 화면 캡처·메타데이터를 구분해 분석 범위를 정하세요.

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.

수집 정책란 무엇인가?

자동 접근 지시와 소유자의 허용 범위를 함께 적용하는 운영 규칙입니다. 이 노트는 수집 정책를 기능 이름보다 입력·변환·출력의 책임으로 읽습니다. 분석 결과를 신뢰하려면 어떤 자료가 들어왔고, 무엇을 확인했으며, 어디까지 결론을 낼 수 있는지 이어져야 합니다.


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?

기술적으로 페이지를 열 수 있어도 수집 권한과 사용 범위가 일치한다고 단정할 수 없습니다. 제한을 실패 점수로 바꾸면 분석도 왜곡됩니다.

설계 원리와 데이터 흐름

자동 수집 지시, 이용 범위와 접근 실패를 구분합니다. 제한이 있으면 승인된 HTML 또는 캡처 자료로 대체 가능한 범위를 확인합니다.

> **수집 정책 확인 → 허용 범위 결정 → 자료 확보·제한 기록**

각 단계는 앞 단계의 성공을 다음 단계의 성과로 바꿔 부르지 않아야 합니다. 자료의 식별자와 기간, 검증 상태를 이어서 기록하면 누락과 오류가 생긴 위치를 찾고 다시 확인할 범위를 정할 수 있습니다.

SAG 아키텍처와의 연결

SAG는 수집 실패·제한을 별도 상태로 남깁니다. 브라우저 기반 수집을 고려하더라도 접근 제한을 우회하는 권한이 생기는 것은 아닙니다.

SAG의 운영 가치는 이 관계를 페이지와 질문, 비교 결과와 개선 작업으로 연결하는 데 있습니다. 고객은 숫자만 읽는 대신 보강할 대상과 판단 근거를 함께 검토할 수 있습니다. 추가 적용이 필요한 패턴은 해당 문단의 범위를 기준으로 읽어야 합니다.

설명용 예제와 판단 기준

설명용으로 robots 지시 때문에 자동 수집을 하지 못했다면 기술 준비도 0점을 부여할 근거가 아닙니다. 미수집 상태를 남기고 승인된 자료의 제출 경로를 제시해야 합니다.

위 예제는 구조와 계산을 설명하기 위한 자료이며 특정 고객사의 실측 성과가 아닙니다. 실제 보고서에는 선택한 기간·대상·관측 조건과 원문 기록이 연결돼야 같은 판단을 다시 확인할 수 있습니다.

실무 검증 체크리스트

흐름 단계	확인할 항목
수집 정책 확인	수집 지시와 계정 권한 구분
허용 범위 결정	제한 사유 보존
자료 확보·제한 기록	대체 자료의 허용 범위 확인

정상 입력뿐 아니라 비어 있는 자료, 중복 자료와 조건이 다른 자료에서도 같은 의미를 유지하는지 확인하세요. 검증 항목을 작업 완료 기준에 연결하면 기능 설명과 실제 운영의 차이를 줄일 수 있습니다.

한계와 적용 시 주의점

RFC 9309의 robots 지시는 접근 인증 자체가 아닙니다. 사이트 권한, 보안 장치와 계약 범위는 별도로 확인해야 합니다.

연구와 공식 문서

- IETF RFC 9309 — 자동 수집에 관한 robots 지시를 정의하며 인증과는 구별합니다.

외부 자료는 위 설계 주제의 배경이며 SAG의 모든 구현이나 고객 성과를 인증하는 자료가 아닙니다. 이 노트의 적용 해석과 설명용 예제는 SAG 운영 구조를 기준으로 정리했습니다. 자료 확인: 2026-10-06.

이어 읽기와 기능 확인

- 관련 아키텍처 노트
- 수집 정책와 연결된 서비스 체험
- 기능별 FAQ
- 도입 범위 상담

이 기술을 이어서 읽는 방법

HTML ZIP·사이트맵 → 정규화 → 페이지 버전 → 근거 기록을 따라갑니다.

- 데이터가 근거가 되는 과정
- 기능 안내 FAQ