

기술 연재 도구가 보고한 숫자를 실제와 대조하는 법 2/4

저희 사이트맵의 날짜를 구글은 무시하고 있었습니다

밖으로 신호를 보내는 일에는 공통된 약점이 하나 있습니다. 그 신호가 상대방에게 어떻게 읽혔는지는 우리 쪽에 안 남는다는 것입니다. 메일을 보내면 보낸편지함에는 남지만 상대가 스팸으로 걸렸는지는 모릅니다. API를 호출하면 200이 돌아오지만 상대 시스템이 그 값을 버렸는지는 모릅니다.

방동걸 (주)웨이스 대표이사

2026년 9월 2일

밖으로 신호를 보내는 일에는 공통된 약점이 하나 있습니다. 그 신호가 상대방에게 어떻게 읽혔는지는 우리 쪽에 안 남는다는 것입니다. VEXPLOR를 만드는 웨이스가 자사 사이트에서 겪은 일이라, 구글이 무엇을 무시했는지 하나씩 적었습니다.

메일을 보내면 보낸편지함에는 남지만 상대가 스팸으로 걸렸는지는 모릅니다. API를 호출하면 200이 돌아오지만 상대 시스템이 그 값을 버렸는지는 모릅니다. 우리 쪽 기록은 전부 정상인데 상대는 아무것도 받지 않은 상태일 수 있고, 그 어긋남은 한참 뒤에 다른 경로로 알게 됩니다.

저희가 그 경로로 알게 된 이야기를 적습니다. 알려 준 것은 저희 검사 장치가 아니라 구글이었습니다.

01

사이트맵이 무엇을 하는 파일인가

검색엔진은 사이트를 통째로 매번 훑지 않습니다. 그래서 사이트들은 사이트맵이라는 목록 파일을 하나 두고 "이 주소들이 있습니다"라고 알려 줍니다.

그 목록에는 주소마다 lastmod라는 날짜가 붙습니다. 이 페이지가 마지막으로 바뀐 날이라는 뜻입니다. 검색엔진은 이 날짜를 보고 다시 읽을지 말지를 정합니다. 안 바뀐 페이지를 매번 다시 읽는 것은 서로에게 낭비이기 때문입니다.

즉 lastmod는 우리가 상대방에게 보내는 신호입니다. 그리고 상대가 그 신호를 어떻게 받아들이는지는 우리 서버 로그에 남지 않습니다.

02

구글이 먼저 알려 줬습니다

시작은 저희 점검이 아니었습니다. 서치콘솔에서 색인되지 않은 페이지가 91건이라는 알림이 왔고, 그것을 전수로 뜯어봤습니다.

결과는 싱겁습니다. **91건 중 실제로 고칠 값이 있는 것은 28건이었습니다.** 나머지는 이미 스스로 풀렸거나(태그 10건) 로그인 페이지·아이콘·피드처럼 애초에 색인 대상이 아닌 것들이었습니다. 28건도 원인이 하나였습니다. 주소 체계를 옮기면서 옛 주소에 리디렉션을 걸지 않은 것.

여기까지는 평범한 정리 작업입니다. 문제는 그다음이었습니다.

03

고쳤는데 구글이 다시 오지 않았습니다

8월 17일에 한글 태그 주소의 404를 고쳤습니다. 그런데 서치콘솔의 마지막 크롤링 기록은 **8월 13~14일에 멈춰 있었습니다.**

고친 것을 상대가 모르고 있었습니다. 그리고 저희는 상대가 모른다는 것을 몰랐습니다.

원인을 찾아 사이트맵 생성 코드를 열었더니, 정적 페이지 8건과 태그 페이지 전량에 `lastmod` 자리에 **지금 시각이 들어가고 있었습니다.** 사이트맵이 요청받을 때마다 만들어지는 구조라, 213건 중 40건 남짓이 **조회할 때마다 방금 바뀌었다고 말하고 있었던 것입니다.**

매번 방금 바뀌었다고 말하는 목록은 아무것도 말하지 않는 것과 같습니다. 검색엔진은 이런 사이트맵의 날짜를 신뢰할 수 없다고 보고 **통째로 무시합니다.** 그래서 진짜로 고친 날에도 그 신호가 전달되지 않았습니다.

저희 쪽에서 보면 모든 것이 정상이었습니다. 사이트맵은 200으로 응답했고, 주소는 전부 살아 있었고, 날짜도 빠짐없이 붙어 있었습니다. 형식은 완벽했고 내용이 무의미했습니다.

04

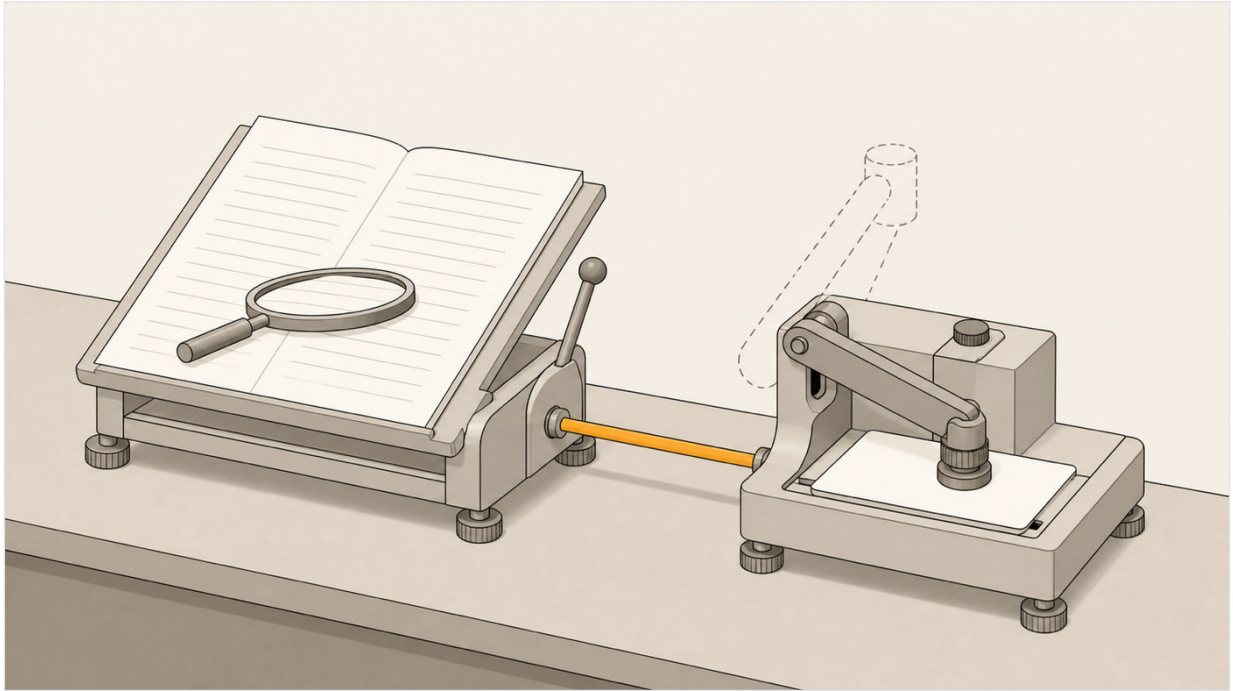
고치다가 더 큰 것이 나왔습니다

날짜를 실제 수정일로 바꿔 배포하고, 라이브 사이트맵을 확인하는 중이었습니다. **홈 화면의 날짜가 아무도 편집하지 않았는데 그날 아침 시각으로 찍혀 있었습니다.**

배포 전과 배포 후의 사이트맵을 각각 저장해 글 62건을 한 줄씩 대조했습니다. 어떤 글의 날짜가 **편집 0건인데 8분 전진해 있었습니다.**

원인은 조회수였습니다. 글을 읽을 때마다 조회수를 1 올리는데, 저희가 쓰는 데이터베이스 도구가 수정하는 모든 요청에 "수정일"을 자동으로 함께 실어 보내고 있었습니다.

그래서 글 상세 62건의 날짜는 줄곧 "마지막으로 고친 날"이 아니라 **마지막으로 누가 읽은 시각**이었습니다. 사이트맵 213건 중 190건 남짓이 조회 때마다 날짜를 바꾸고 있었다는 뜻입니다.



읽기만 했는데 고친 날이 바뀌었습니다. 읽는 쪽과 기록하는 쪽이 원래 물려 있으면 안 되는 것이었습니다

이건 이번에 만든 문제가 아니라 **원래 있던 문제**입니다. 다만 정적 페이지의 날짜를 고치면서 값이 목록 전체로 넓어졌고, 그 과정에서 드러났습니다.

05

고친 방법

세 가지를 했습니다.

- **조회가 수정일을 밀지 않게 했습니다.** 조회수만 올리는 쿼리를 따로 써서, 읽는 행위가 "고쳤다"는 신호를 만들지 않게 했습니다.
- **날짜를 못 구하면 아예 신지 않기로 했습니다.** 데이터베이스 조회가 실패했을 때 종전에는 현재 시각을 대신 넣었습니다. 지금은 그 자리를 비웁니다. 거짓 날짜보다 없는 편이 낫습니다.
- **매일 도는 검사기를 만들었습니다.** 사이트맵을 두 번 조회해 날짜가 그대로인지 봅니다. 값이 달라지면 다시 그 상태로 돌아간 것입니다.

세 번째가 이 사건의 답입니다. 앞의 둘은 지금 난 문제를 고치는 일이고, 세 번째만 같은 문제가 다시 났을 때 알려 줍니다.

06

이 사건에서 남는 것

- 우리 쪽 검사는 형식을 봅니다. 응답 코드, 필드 존재, 스키마 적합. 이번 사이트맵은 그 검사를 전부 통과하는 상태로 무의미했습니다.
- 신호는 보내는 쪽이 아니라 받는 쪽 기준으로 재야 합니다. "날짜를 보냈는가"가 아니라 "그 날짜가 상대방에게 쓸모 있는가"가 질문이었습니다.
- 밖에서 온 알림은 대부분 그 알림의 내용이 본론이 아닙니다. 404 알림에서 시작했는데 실제로 고칠 것은 404가 아니었습니다. 알림은 들여다보라는 신호였을 뿐입니다.

그리고 하나가 남습니다. 이번에는 바깥이 알려 줬고, 저희는 그 지적을 받아 저희 데이터를 열어 봤습니다. 그런데 그때 저희가 근거로 삼은 그 데이터는 원본이었습니까.

같은 화면을 3개월 동안 보면서 엉뚱한 자리를 고치고 있던 이야기를 다음 회차에 적겠습니다.

WACE(웨이스)는 공장이 스스로 판단하고 멈추지 않게 하는 제조 AI 운영체제 VEXPLOR를 만드는 회사다.



원문 blog.vexplor.com/post/our-sitemap-dates-were-ignored

발행 2026년 9월 2일 · 방동걸 (주)웨이스 대표이사

회사 주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 ·
사업자등록번호 439-87-02852

주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동)
1332·1333·1334·1335호 (영천동, 실리콘엘리동탄)

연락처 전화 070-4730-0629 · contact@wace.me

홈페이지 wace.me · linkedin.com/company/wace-inc



원문



홈페이지



링크드인