

## 비즈니스

# 저희 블로그 태그 226개 중 163개가 글 한 편짜리였습니다

저희 기술블로그의 태그를 전부 세어 봤습니다. 글 106편에 태그가 226개였고, 그중 163개(72%)가 글 한 편에만 붙어 있었습니다. 태그 하나에 글 한 편. 독자가 그 태그를 누르면 글 한 편이 나옵니다. 방금 읽고 있던 그 글입니다.

방동걸 (주)웨이스 대표이사

2026년 9월 7일

저희 기술블로그의 태그를 전부 세어 봤습니다. 글 106편에 태그가 226개였고, 그중 163개(72%)가 글 한 편에만 붙어 있었습니다. VEXPLOR 기술블로그를 운영하는 웨이스가 그 숫자를 왜 문제로 봤는지부터 차례로 적습니다.

태그 하나에 글 한 편. 독자가 그 태그를 누르면 글 한 편이 나옵니다. 방금 읽고 있던 그 글입니다.

## 01

## 세어 보게 된 계기는 구글이 준 알림이었습니다

검색엔진 색인 리포트에 "발견했지만 아직 수집하지 않음"이 20건 잡혔습니다. 확인해 보니 20건 전부 살아 있고(200), 색인 차단도 없고, 사이트맵에도 있었습니다. 우리 쪽에 막아 둔 것은 없었습니다.

그런데 그 20건의 구성이 눈에 걸렸습니다.

| 구분     | 건수 | 예               |
|--------|----|-----------------|
| 태그 페이지 | 16 | 검색 · 무인공장 · seo |
| 고정 페이지 | 3  | 소개 · 개인정보 · 약관  |
| 글      | 1  | 사흘 전 발행분        |

16건이 태그였습니다. 검색엔진이 수집을 미룬 것과 우리가 가장 많이 만든 것이 같은 종류였습니다. 그래서 태그 전체를 세게 됐습니다.

02

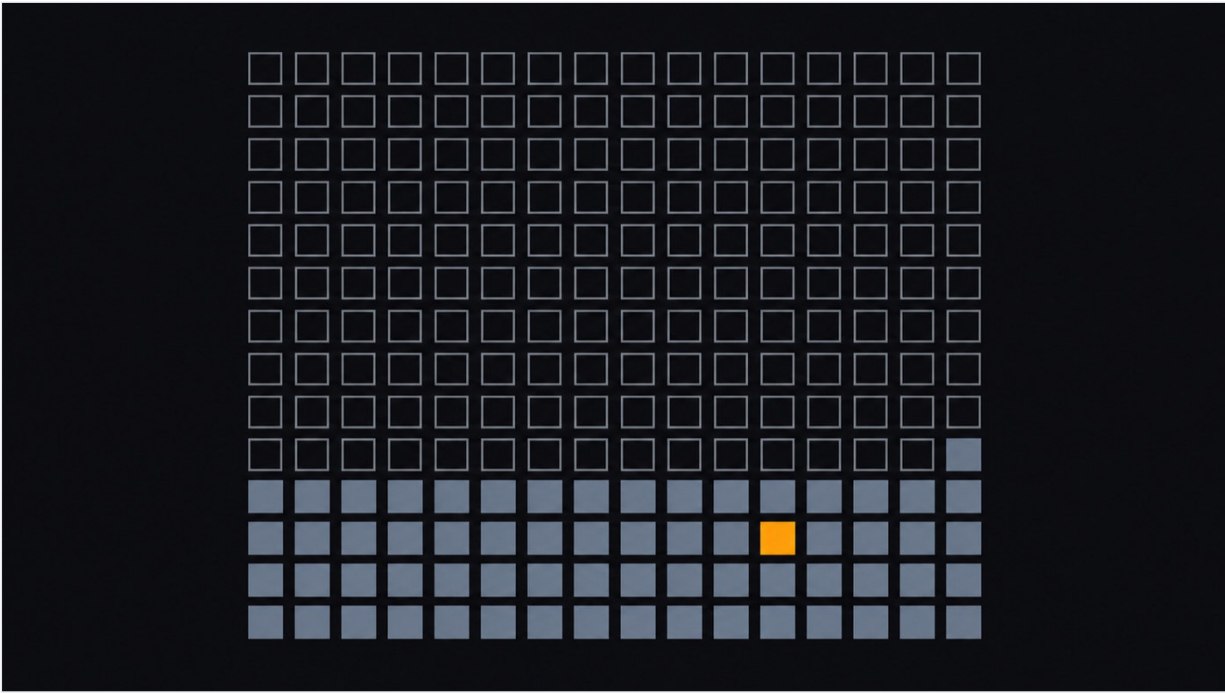
태그 226개의 분포

226개 전건에 걸린 글 목록까지 수집했습니다. 결과입니다.

| 태그에 달린 글 | 태그 수 | 비율  |
|----------|------|-----|
| 1편       | 163  | 72% |
| 2편       | 25   | 11% |
| 3~5편     | 26   | 12% |
| 6편 이상    | 12   | 5%  |

사이트맵은 347건이고 그중 태그가 226건입니다. 사이트맵의 47%가 글 한 편만 담은 페이지라는 뜻입니다. 글 한 편을 쓰면 태그 페이지가 평균 2.1장 생겼습니다.

발행할 때 태그를 5개쯤 답니다. 그 5개를 매번 새로 지으면 이렇게 됩니다. 기존 태그를 찾아보는 것보다 지금 떠오른 말을 적는 게 빠르니까요.



태그 226개 가운데 163개가 글 한 편만 담고 있습니다. 빈 칸이 그만큼입니다.

## 03

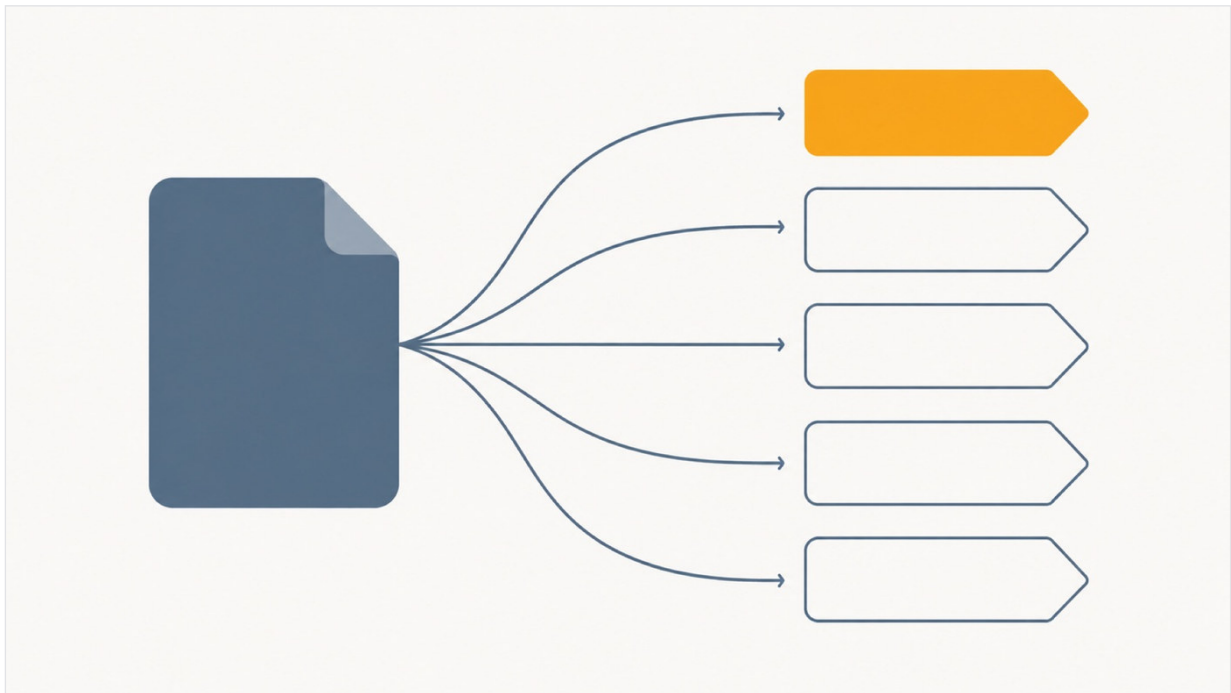
## 어떻게 썼는지

방법은 단순합니다. 사이트맵에서 태그 주소를 전부 뽑고, 태그 페이지를 하나씩 열어 그 안에 글 링크가 몇 개인지 셉니다. 226번 요청하면 끝납니다.

- **목록은 사이트맵에서** — 태그 목록 페이지를 긁는 것보다 정확합니다. 사이트맵은 우리가 색인해 달라고 내미는 것이라, 거기 있는 것이 곧 검색엔진이 보는 목록입니다.
- **글 수는 태그 페이지에서** — 페이지 안의 `/post/` 링크를 중복 제거해 셉니다. 목록이 여러 쪽으로 나뉘면 쪽수까지 따라가야 하는데, 저희는 가장 많은 태그가 20편이라 한 쪽에 들어왔습니다.
- **8개씩 동시에** — 226번을 순서대로 하면 몇 분 걸립니다. 우리 서버라 부담이 없어 병렬로 돌렸습니다.

여기서 한 번 크게 틀릴 뻔했습니다. 처음 뽑은 목록으로 대조했더니 **20건 전부 "사이트맵에 없음"으로 나왔는데, 실제로는 20건 전부 있었습니다**. 주소를 뽑는 명령 한 줄이 조용히 어긋나 있었고 오류는 나지 않았습니다. 그 이야기는 [검사 결과가 0건일 때 없는 것과 고장 난 것을 어떻게 가르는지에](#) 따로 적었습니다.

그래서 이 글의 숫자들은 두 번 다른 방식으로 세어 맞춘 값입니다. 한 번은 태그별로, 한 번은 글별로 셉고 합계가 일치했습니다.



글 한 편을 쓰면 태그 페이지가 평균 2.1장 생깁니다.

## 04

## 낱말 뜻으로 짝지으려 하자 틀렸습니다

처음에는 163개를 하나씩 짝지어 합칠 생각이었습니다. 이름이 비슷한 것끼리 묶으면 될 것 같았습니다.

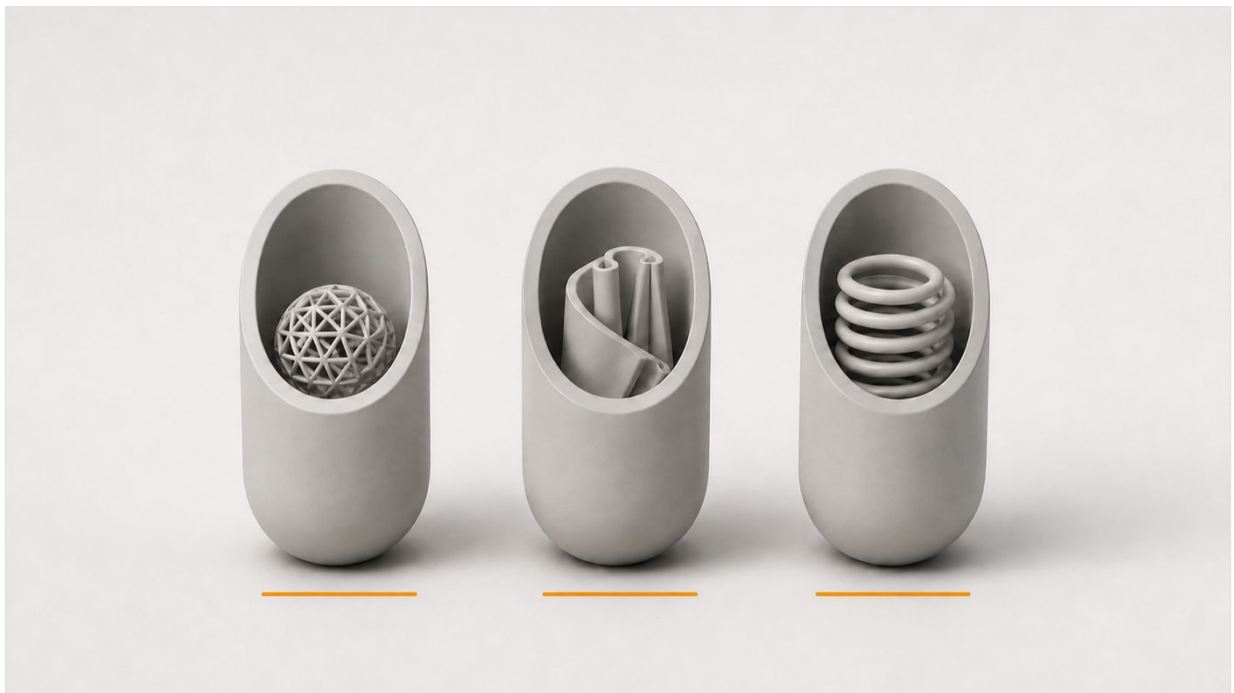
두 번 틀렸습니다.

**첫 번째.** 검색과 ai검색을 합치려 했습니다. 이름이 겹치니까요. 그런데 붙은 글을 열어 보니 검색은 표본을 원본으로 착각한 이야기였고, ai검색은 AI가 우리 사이트를 읽는 이야기였습니다. 다른 주제입니다.

**두 번째.** 데이터모델링 · 데이터설계 · 데이터아키텍처 세 개가 한 무리로 보였습니다. 열어 보니 각각 온톨로지의 시간 표현, 개인정보 최소 수집, 검색 구조 비교였습니다. 셋 다 다른 글입니다.

기계로 뽑은 어간 묶음에는 edisonawards · isa-95 · microsoft 가 "is"를 공유한다는 이유로 한 무리가 된 것도 있었습니다.

**이름이 닮은 것과 뜻이 같은 것은 다릅니다.** 어간 겹침은 후보를 찾는 데만 쓰고, 판정은 글을 열어서 했습니다.



겉이 같아 보이는 세 태그를 열어 보니 안이 전부 달랐습니다.

05

## 뜻이 통하는 기준은 "한 글이 만든 뭉치"였습니다

각도를 바꿔서, 태그가 아니라 글을 기준으로 봤습니다. 한 글에 1편짜리 태그가 몇 개나 붙었는지.

- 1편짜리를 2개 이상 만든 글 — 50편
- 그 글들이 만든 태그 — 128개
- 163개에서 차지하는 비중 — 79%

50편이 163개 중 128개를 만들었습니다. 그리고 그 50편 중 47편은 이미 2편 이상 붙은 태그도 함께 달고 있었습니다. 1편짜리를 하나도 안 붙였어도 그 글은 다른 태그로 찾을 수 있었다는 뜻입니다.

가장 심한 글은 태그 5개가 전부 1편짜리였습니다. 그 글 하나가 색인 페이지를 5장 만든 셈입니다.

이 기준은 추측이 필요 없습니다. "이 태그는 그 글 하나만 가리키는가"만 보면 되니까요.

06

## 163개가 전부 문제인 것은 아니었습니다

세는 김에 하나 더 갈랐습니다. 1편짜리 중 48개는 1편이 맞았습니다.

- **고유명사** — 협력사 이름, 정부 부처, 대회 이름. 뉴스 글에 한 번 나오고 마는 게 정상입니다. 합치면 그 회사 이름으로 찾아온 사람이 글을 못 찾습니다.
- **규격 번호** — isa-95 · iec63278 · ksx9001 같은 것들. 이건 검색어 그 자체라 다른 말로 바꾸면 값이 없어집니다.
- **그 글에만 있는 주제** — 충돌 회피, 경로 계획, 회귀 테스트. 아직 한 편뿐이지 언젠가 두 번째 글이 붙을 자리입니다.

그러니까 163개는 "줄여야 할 부채 163개"가 아니라 **줄일 것 115개와 남길 것 48개**였습니다. 세어 보기 전에는 이 구분이 없었고, 만약 숫자만 보고 일괄 정리를 돌렸으면 뉴스 글의 회사 이름을 지웠을 겁니다.

숫자를 줄이는 것 자체가 목표가 되면 이런 걸 놓칩니다.

07

## 합칠 수 있는데 합치지 않았습니다

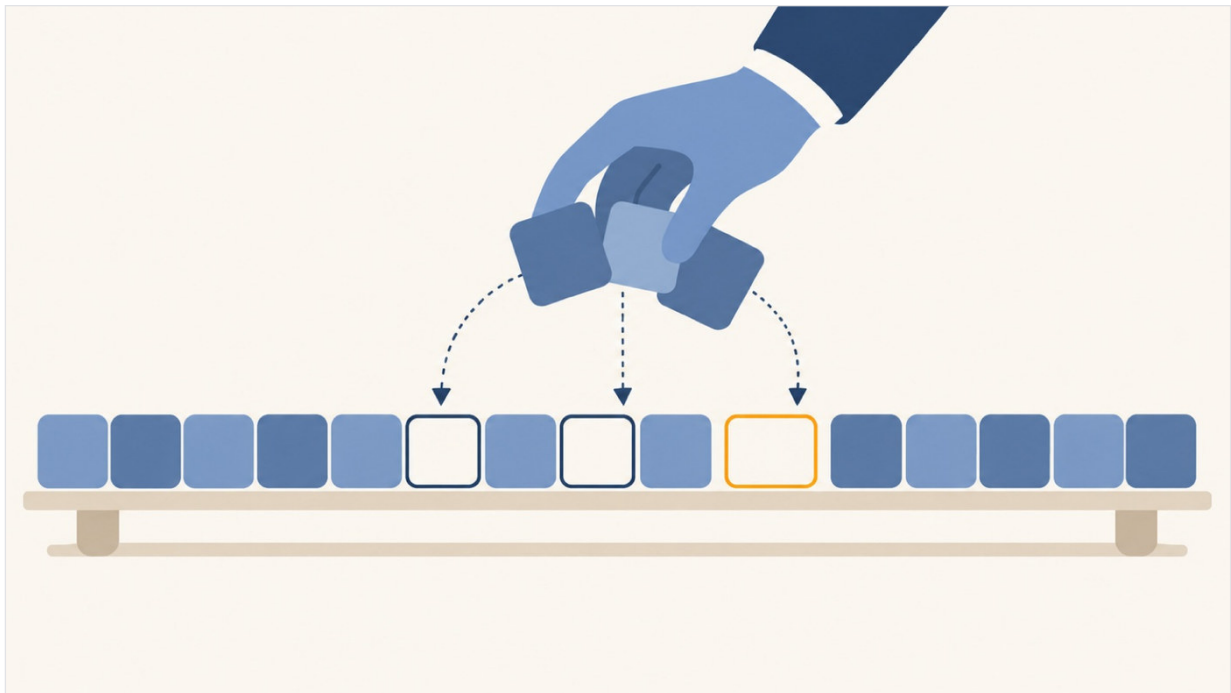
계산해 보면 최대 125개를 줄일 수 있었습니다. 태그 226개가 101개가 됩니다.

하지 않았습니다.

태그를 떼면 그 태그 페이지가 사라집니다. 사라진 주소는 404가 됩니다. 저희는 몇 주 전에 404를 줄이려고 옛 주소에 리디렉션을 붙이는 작업을 했는데, 태그를 정리하면 125장을 새로 만드는 셈입니다. 그걸 막으려면 125개의 리디렉션을 다시 붙여야 하고요.

그리고 지금 아프지 않습니다. 사이트맵 347건 중 색인된 것이 333건, **\*\*96%\*\***입니다. 수집이 밀린 20건도 2주 전 50건에서 내려오는 중입니다.

숫자가 커 보인다고 손대는 것과, 그 숫자가 실제로 무엇을 깎고 있는지 확인하고 손대는 것은 다릅니다. 지금은 후자의 근거가 없습니다.



태그를 떼면 그 자리가 빈 주소로 남습니다. 되돌리려면 하나씩 다시 이어야 합니다.

08

## 대신 늘어나는 속도를 멈췄습니다

문제는 현재 상태가 아니라 방향입니다. 글이 300편이 되면 태그는 600개를 넘고, 그때는 얇은 페이지가 실제로 걸리기 시작합니다.

그래서 과거는 두고 앞으로만 바꿨습니다.

- 새 글의 태그는 기존 태그에서 고릅니다. 새 태그는 앞으로 2편 이상 쓸 것이 분명할 때만 만듭니다.
- 정본표를 만들었습니다. 같은 뜻인데 다르게 적힌 20쌍을 정리했습니다. 에이전트와 ai에이전트, 생산관리시스템과 제조실행시스템과 mes 같은 것들입니다. 한 글에 뜻이 같은 태그가 셋 붙은 경우도 있었습니다.

- **발행 전에 한 번 확인합니다.** 스크립트에 태그를 넣으면 기존인지 새 것인지, 기존이면 지금 글이 몇 편인지 알려 줍니다.

```
$ python3 태그_점검.py 제조ai특화 하네스 검증설계 진단
[교체] 제조ai특화 → 제조ai 같은 뜻의 늘려 쓴 표기
[교체] 하네스 → 하니스 같은 말의 표기 흔들림
[기존] 검증설계 지금 글 11편
[교체] 진단 → 성속도진단 이 블로그의 진단은 성속도 진단이다
```

막는 게 아니라 묻는 형태입니다. 새 주제를 시작하는 자리라면 새 태그가 맞으니까요.



정분표는 왼쪽 대신 오른쪽을 쓰라고 알려 줍니다. 과거에 붙은 태그는 그대로 둡니다.

09

## 방향이 반대인 규칙을 옆에 두게 됐습니다

여기서 조심할 게 하나 생겼습니다. 저희는 네이버 블로그도 운영하는데, 거기서는 태그 규칙이 정반대입니다.

| 구분    | 자체 블로그    | 네이버     |
|-------|-----------|---------|
| 원하는 것 | 겹치게       | 겹침 0    |
| 이유    | 얇은 페이지 방지 | 자기잠식 방지 |

네이버는 태그가 검색 노출 단위라, 우리 글끼리 같은 태그를 쓰면 서로 순위를 나눠 갖습니다. 그래서 매 글이 새 태그입니다. 자체 블로그는 태그마다 색인되는 페이지가 한 장씩 생기니 반대입니다.

같은 낱말 "태그"인데 규칙이 거꾸로입니다. 그래서 네이버 태그 원장 맨 위에 이 대조표를 붙여 뒀습니다. 규칙을 정보 문서에만 적으면, 반대 규칙을 쓰는 사람은 그것을 볼 일이 없습니다.

10

## 남는 이야기

이 작업에서 실제로 얻은 것은 태그 정리가 아니라 판정 기준이었습니다.

숫자를 발견하면 고치고 싶어집니다. 163개는 큰 숫자니까요. 그런데 고치는 쪽에도 비용이 있고, 이번에는 그 비용이 리디렉션 125개와 배포 한 번이었습니다. 얻는 것은 "얇은 페이지가 준다"인데, 지금 그게 무엇을 깔고 있는지는 측정되지 않았습니다.

그래서 다시 볼 조건을 적어 두고 닫았습니다. 태그가 300개를 넘거나, 색인 커버리지가 90% 아래로 내려가거나, 수집 대기가 다시 40건대로 오를 때. 그 셋에 해당하지 않으면 이 안건을 다시 열지 않습니다.

색인이 왜 밀리는지 궁금해서 들여다본 이야기는 [사이트맵 낱자를 구글이 무시하고 있었던 건에도](#) 적었습니다. 그때도 이번에도, 문제를 먼저 알려 준 것은 저희 계기판이 아니라 검색엔진이었습니다.

여러분의 블로그는 태그가 몇 개입니까? 그중 글 한 편짜리는 몇 개일까요?

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

원문 [blog.vexplor.com/post/most-of-our-tags-pointed-to-one-post](https://blog.vexplor.com/post/most-of-our-tags-pointed-to-one-post)

발행 2026년 9월 7일 · 방동걸 (주)웨이스 대표이사

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

주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동)  
1702-1706호 (영천동, 실리콘밸리동탄)

연락처 전화 070-4730-0629 · [contact@wace.me](mailto:contact@wace.me)

홈페이지 [wace.me](https://wace.me) · [linkedin.com/company/wace-inc](https://linkedin.com/company/wace-inc)



원문



홈페이지



링크드인