

Primary vs. Supplemental Index: Checking Your Pages

When a fair chunk of your revenue depends on organic traffic, the difference between a page living in the **primary index** and one dumped into the supplemental bucket is the difference between money and silence. The phrase Primary vs. Supplemental Index: Checking Your Pages sums up the single most underrated hygiene check most site owners skip — and then wonder why half their content never sees a click.

Google hasn't used the term "supplemental index" publicly for a decade, yet the concept still lives on in log files and in how the main index prioritises what to show. Checking which pages actually get treated as first-class citizens of the main index isn't a theoretical exercise. It's a practical skill you develop after staring at too many rank-tracking dashboards that show zero for articles you were sure would perform.

Over the years, I've watched perfectly reasonable pages get labelled "Crawled - currently not indexed" while spammy junk sailed into the primary index. The gap between those two outcomes often boils down to signals a URL inspector can't show you directly — but that the tools we'll cover expose if you push them hard enough.

What "Primary Index" Actually Means (and Where the Supplemental Concept Came From)

Think of the main index as Google's fast-access store: hundreds of billions of pages that get served for queries because they pass a combined quality, freshness, and authority threshold. The supplemental index, when it was announced in 2004, was a separate, slower-fetched pool for pages with thinner content, duplicate-ish structures, or credibility issues. Those pages could still rank, but only for extremely narrow, low-competition queries, and they rarely carried the same visibility.

Today Google doesn't expose a separate "supplemental" label. Instead, a page either enters the main index or shows up in Search Console as "Discovered - currently not indexed", "Crawled - currently not indexed", or simply gets ignored. It's the same logic, rebranded: the indexer still discriminates, just silently.

Rule of thumb: If a page isn't generating impressions for any of its target queries after three weeks and it's not blocked, it's not sitting in the primary index — or it's been deprioritised so aggressively it might as well be.

A 2023 analysis by an industry data provider estimated that roughly 20–30% of pages on mid-sized e-commerce sites fall into this grey zone — crawled but never allowed into the main result set. That's a lot of wasted crawl budget.

The Tools You Need to Verify Index Status

You can't use a single method and trust the answer. Each one gives you a part of the puzzle. Here's a five-point checklist of the instruments you'll actually rely on, ranked by reliability:

- **Google Search Console URL Inspection tool** – canonical source of truth for live index status, showing “URL is on Google” with the last crawl date. [Official docs](#).
- **Index Coverage report** – reveals bulk classifications: “Submitted and indexed”, “Crawled – currently not indexed”, “Discovered”, etc.
- **site: operator combined with inurl:** – quick manual snapshot, but dangerously misleading for pages that rank only for very obscure stems.
- **Google Indexing API** (for job posting/broadcast page types) – lets you request indexing and query status programmatically with a project quota of 2000 requests/day.
- **Third-party bulk index checkers** like [SpeedyIndex's Google Index Checker API](#) — capable of passing thousands of URLs through a verification pipeline in minutes, often returning structured JSON with index status, snippet, and canonical info.

Step-by-Step: Diagnosing Whether a Page is in the Main Index

Here's the exact sequence I follow when a client swears a page is live but the data says no. No vague recommendations.

1. Open GSC and paste the exact URL into the inspection bar. Look at the status line. If it says “URL is on Google”, you're in the primary index. If it says “URL is not on Google”, click “Request Indexing” and note the covering check — that's a signal, not a

fix.

2. Expand the “Coverage” section inside the inspection panel. If the page was crawled but is “Excluded” for any reason other than a deliberate `noindex`, you’ve got a supplemental candidate. Common flags: “Duplicate without user-selected canonical”, “Page with redirect” when none exists, “Crawled - currently not indexed”.

3. Cross-check with a direct search. Use `site:example.com “exact title of the page”` or craft a query that matches a unique fragment. If the page appears, it’s indexed — but many supplemental-class pages still show up for extremely long-tail queries that nobody actually types. So this step confirms presence, not primary-index quality.

4. Look at the Coverage report for the whole site, filtered by “Excluded”. Sort by count descending. That’s your supplemental bucket, practically speaking. In one audit, a client had 14,000 URLs under “Crawled - currently not indexed” — all near-identical product variants with a single parameter swap. That’s a textbook case.

5. For bulk verification, skip the manual clicking and run an API pass. The following flowchart maps the decision path when you need to classify thousands of URLs at once:

```
```mermaid
graph LR
 A[Submit URL batch] --> B{getIndexStatus}
 B -- "indexed: true" --> C[Log as primary]
 B -- "indexed: false" --> D{Reason?}
 D -- "crawled, not indexed" --> E[Flag for content review]
 D -- "discovered, not indexed" --> F[Check internal link strength]
 D -- "blocked by robots" --> G[Fix robots.txt or meta]
```
```

When a Page Sits in Supplemental Territory - What Triggers It

People get stuck on “duplicate content” as the only explanation. Reality is messier. These are the patterns I’ve seen trigger index demotion across dozens of sites — and a few myths that keep floating around.

Myth: Using a `noindex` tag on thin pages fixes supplemental problems.

Reality: That only removes them from the main index entirely. It doesn’t improve the indexing health of nearby pages — but it can free up crawl budget for pages that matter.

Myth: If a page appears when you search `site:example.com/url`, it's in the primary index.

Reality: The site operator returns pages from all tiers, including those that would never rank for a competitive head term. The only trustworthy signal is the Search Console "URL is on Google" combined with tracking organic impressions.

Myth: Submitting URLs via the URL Inspection tool or sitemap guarantees primary index inclusion.

Reality: It adds them to a crawl queue. Google still decides which pages earn a spot in the main result set based on signals. Sitemaps don't override quality evaluation.

Real trigger examples: e-commerce pages where only the color attribute changes and every variant generates its own title tag identical to the parent. Internal search result pages that got accidentally left open and crawled. Paginated URLs with no canonical pointing back to the view-all page. All of these breed massive excluded-URL populations.

[Index Your Backlinks in Record Time →](#)

Real-World Check: Bulk Index Verification with Scripts

When you need to audit 10,000 URLs and separate the primary-index pages from the rest, manual inspections won't cut it. You run a script. Below is a working Python loop that hits a bulk index-checking API — I've used a similar pattern with the SpeedyIndex endpoints — to retrieve a clean yes/no per URL along with canonical and snippet data.

```

```python
import requests
import json
import time
API_KEY = "YOUR_SPEEDYINDEX_API_KEY"
ENDPOINT = "https://en.speedyindex.com/api/v2/check"
urls = ["https://yourdomain.com/primary-product-page", "https://yourdomain.com/thin-variant-page", "https://yourdomain.com/outdated-blog-post"]
headers = { "Authorization": f"Bearer {API_KEY}", "Content-Type": "application/json" }
Prepare payload — max 1000 URLs per request per docs
payload = {"urls": urls}
response = requests.post(ENDPOINT, json=payload, headers=headers, timeout=60)
data = response.json()
for result in data.get("results", []):
 url = result["url"]
 indexed = result["indexed"]
 # True if in primary index
 snippet = result.get("snippet", "")
 canonical = result.get("canonical", "")
 # Log or flag based on status if not indexed:
 print(f"NOT IN MAIN INDEX: {url} (snippet: {snippet})")
 else:
 print(f"MAIN INDEX: {url}")
 time.sleep(1)
Respect rate limits
```

```

Expect a `429` if you push more than 60 requests per minute on some plans, so proper throttling is non-optional. The JSON response usually includes a `reason` field when a page isn't indexed, which maps to the same GSC coverage labels.

:::warning Bulk index checkers rely on Google's own signals plus proprietary sampling. They're not 100% synchronous with the live index at this instant — a 5-minute lag is common. :::

You can also query Google's own Indexing API for a smaller set of URLs if your site falls under the job posting/broadcast use case. But for general web pages, Google's API will only return a `PERMISSION_DENIED` unless you are using it for one of the approved content types. That's why third-party solutions exist.

Common Mistakes That Make You Think You're Indexed (But You're Not)

1. The “site: domain” trap. You see the page in search results, but it's actually the `https://` version when your canonical points to `https://`, or vice versa, and the displayed URL is from a cache layer. The inspection tool shows the real story.

2. Assuming “Crawled - currently not indexed” will resolve on its own. In most cases, it never does without internal link re-balancing or content trimming. I've tracked clusters of URLs in this state for months — zero movement.

3. Checking only the homepage and top category pages. The pages most likely to be stuck in supplemental limbo are deep-level product filters, old blog posts, press

releases, and tag pages. Skip those and you're blind to where the real waste lives.

4. Forgetting about mobile-first indexing fallout. A desktop version that gets indexed while the mobile version is blocked or missing structured data can create a split index situation, effectively pushing the URL toward the lower tier.

Quick Questions People Always Ask

Q: Is there still a literal “supplemental index” inside Google?

No. The term died officially around 2007, but the functional separation remains: pages that meet the quality threshold enter the main serving system; the rest are inventoried but rarely served for normal queries.

Q: Does a page that returns “Crawled - currently not indexed” take up crawl budget?

Yes, it wastes crawl budget because Googlebot already spent the resources to fetch it. The absence of indexing signals is the problem, not the crawl itself.

Q: Can I force a page from supplemental into the primary index with a simple re-submit?

No. The inspection “Request Indexing” button is a crawl request, not a quality re-evaluation switch. You'll need to improve content, consolidate duplicates with 301s, or remove thin pages altogether.

Q: Should I block thin pages with robots.txt to save crawl capacity?

Only if you're absolutely certain they'll never need to appear in search. A `noindex` with a `follow` directive is safer because it allows Google to see links flowing through the page while keeping the URL out of the index.

Q: How fast do bulk index checkers update their data?

Many refresh daily from Google's live index through sampling; real-time is not guaranteed. For urgent verification, use GSC's inspection tool in small batches.

What to Do After You Know Your Index Status

Stop overhauling your site architecture. Take the list of pages that are not in the primary index and apply one of two treatments: fix them or kill them. A page that accumulated

zero impressions over 90 days and sits in the excluded bucket has already proven its worth. Remove it, 301 it to a stronger parent if there's any link equity, and let the crawl budget flow elsewhere.

For pages you want to rescue — perhaps a product page where the copy is two sentences — add at least 300 words of differentiated non-boilerplate content, internal links from real category pages, and a `self-referencing canonical` tag. Submit the URL through GSC one more time, then watch the Coverage report for the next 10 days. If it flips to “Submitted and indexed”, you've won.

Checking the divide between primary and supplemental isn't something you do once. It's a monthly audit that keeps your site's footprint clean and your server's crawl allowance pointed at the URLs that actually earn traffic.

Further Reading

1. IndexNow. "Protocol Overview." indexnow.org
2. Bing Webmaster. "Submit Sitemaps." bing.com/webmasters
3. Google Search Central. "Robots.txt Introduction." developers.google.com