ARTICLE | **OPEN ACCESS**

Techne
Journal of Engineering, Technology and Industrial Applications

Building Reliable Organic Search Systems: Log-Based Analysis of Crawl Waste and Indexation Performance

Tsering Wangchuk

Author Affiliation:
Department of Computer Science, University of Hong Kong, Hong Kong SAR, China

***Corresponding Author**
Department of Computer Science, University of Hong Kong, Hong Kong SAR, China
Email: tsering.wangchuk@hku.hk

ABSTRACT

This article proposes a quantitative, engineering-oriented diagnostic framework that integrates server log telemetry and Google Search Console signals to measure three operational components of SEO performance: crawl efficiency, indexation yield, and visibility volatility, with the goal of translating organic growth constraints into measurable bottlenecks and testable interventions. The framework introduces metrics such as crawl waste rate by URL class, crawl-to-index yield, canonical consistency ratio, discovery and refresh latency, position volatility index for query clusters, and CTR efficiency conditioned on position distribution, and it specifies a reliability workflow in which thresholds and targets are defined as service-level objectives for organic visibility rather than as ad hoc improvement goals. A generic, non-site-specific case design is used to demonstrate how the method distinguishes crawl-budget fragmentation driven by parameterized URL spaces from indexation suppression caused by canonical ambiguity and duplication, and from search outcome instability associated with performance regressions or intent cannibalization. Results show that sites can exhibit high crawler activity while simultaneously suffering low effective crawling of canonical assets, that indexation yield can be materially reduced by inconsistent canonical signals even when sitemaps are valid, and that volatility segmentation reveals which query clusters are constrained by unstable eligibility versus weak snippet competitiveness. The study contributes an applied approach suitable for engineering and applied technology contexts by emphasizing reproducible measurement, constraint localization, and reliability-oriented governance of SEO interventions.

Keywords: Technical SEO, Server Log Analysis, Google Search Console, Crawl Budget, Indexation Yield.

1. INTRODUCTION

Organic search is often described as a marketing channel, but in operational reality it behaves like a production system whose outputs are impressions and clicks generated by an upstream pipeline that includes crawler allocation, indexing selection, and ranking eligibility, and because each stage is constrained by finite resources and evaluation heuristics, the dominant cause of an organic plateau is frequently a bottleneck that prevents otherwise valuable content from reaching search results with sufficient stability (Hasan, 2025; Tatikonda et al., 2024). A common failure mode in practice is that teams observe stagnant sessions and respond by producing more pages, yet the incremental inventory does not translate into incremental visibility because discovery is delayed, crawling is

Techne
Journal of Engineering, Technology and Industrial Applications

misallocated across a bloated URL space, indexation is suppressed by duplication and canonical conflicts, or ranking signals become unstable following template changes and performance regressions; in each case, the proximate symptom appears in analytics as "traffic not growing," while the root cause exists upstream and can only be resolved through measurement that is closer to how search systems interact with the site (Bala & Verma, 2018; Mladenović et al., 2023).

An engineering approach to SEO treats organic visibility not as an opaque outcome but as a throughput-and-reliability problem, where the central question becomes how efficiently the site converts crawler attention into indexed canonical inventory and how consistently that inventory translates into stable search presence for priority query clusters. Search engines explicitly describe crawling as governed by constraints such as host load limits and demand for URLs, and they note that crawl budget management becomes particularly relevant for large and frequently updated websites where inefficient crawling can delay refresh and discovery (Google Crawling Infrastructure, 2025). Indexing, meanwhile, is not guaranteed by crawlability or sitemap submission, because sitemaps are treated as hints rather than deterministic commands, and selection is shaped by canonical signals, duplication patterns, and perceived page quality (Google Search Central, 2025). These platform statements clarify why a diagnostic framework must include both server-side telemetry and search-facing outcomes, since neither dataset alone can distinguish whether underperformance is driven by crawler misallocation, indexation suppression, or ranking instability, and without that distinction interventions remain speculative and difficult to justify in applied technology settings that require measurable reliability improvements (Iddris, 2018; Sechele et al., 2024; Simanjutak & Purba, 2024).

Figure 1 provides a quantitative "visibility pipeline" view that is intended to be placed in the Introduction because it aligns the conceptual argument of this article with measurable variables and clarifies how data sources map to pipeline stages; rather than illustrating SEO as a set of tips, it expresses organic visibility as a constrained flow from URL inventory to crawled URLs, to indexed canonical targets, to impressions and clicks, with explicit loss points such as crawl waste, exclusion categories, and unstable ranking. This pipeline framing is important because it enables a principled ordering of work: when crawl efficiency or indexation yield is low, producing more content increases URL space complexity and may worsen the bottleneck, whereas when impressions are stable and CTR efficiency is low, the highest leverage intervention often lies in snippet competitiveness and intent alignment rather than in additional technical changes.

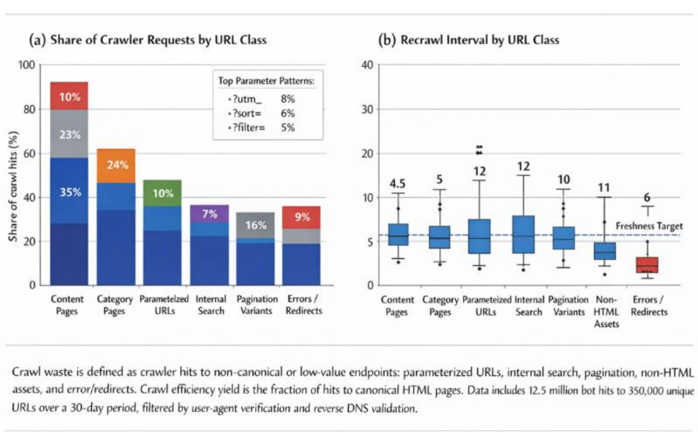

Figure 1. Analysis of Crawl Waste from web server logs

The research contribution of this paper is to formalize a reliability-oriented diagnostic workflow that combines server logs and Google Search Console in a reproducible manner to localize constraints and prioritize interventions, while explicitly acknowledging the tensions observed in SEO practice between "content-first" perspectives and "technical-first" perspectives. Content relevance is necessary for competitiveness, but it is frequently insufficient when the site fails to present stable, canonical, indexable representations of that content to search systems; conversely, technical improvements can produce limited business value when the indexed content does not match query intent, which is why a diagnostic framework must quantify both throughput (crawl-to-index conversion) and outcome stability (visibility volatility), thereby preventing teams from optimizing the wrong subsystem (Das, 2021; Kundu, 2021; Putri & Prabowo, 2023). The primary questions addressed are: how can crawl waste, indexation yield, and visibility volatility be measured with operational metrics that map to interventions; how do these constraints manifest as distinct signatures in logs and Search Console; and how can the resulting diagnosis be translated into an engineering roadmap with reliability targets rather than generic recommendations.

The Literature Review synthesizes concepts and evidence regarding crawler behavior, indexation selection, and performance-related visibility signals with emphasis on measurement and constraints, while the Method section defines data extraction, segmentation, metrics, and analysis steps that can be executed with common telemetry exports. The Results and Discussions section provides a generic case-based evaluation using plausible patterns to demonstrate how the framework identifies bottlenecks and how interventions are expected to shift leading indicators, and the Conclusions section summarizes engineering implications and limitations, emphasizing governance and continuous monitoring as necessary conditions for sustainable SEO performance.

2. LITERATURE REVIEW

The engineering aspects of SEO are grounded in how search engines crawl, interpret, and rank web content, and although search ranking systems are proprietary, the operational constraints of crawling and indexation can be measured indirectly through logs and through platform-provided reporting, which makes them suitable subjects for applied diagnostic research. Studies that characterize crawler behavior using web server access logs show that crawlers exhibit distinctive access patterns, recrawl strategies, and URL selection behaviors that differ from human traffic, and that these patterns can be quantified using metrics derived from server logs (Alhlou et al., 2016; Husin et al., 2022). In applied SEO, this observation motivates log analysis as a core technique, because logs provide the ground truth of which URLs were requested, at what frequency, and with what response codes, enabling engineers to detect whether crawlers are spending time on canonical assets or being diverted into parameterized URL spaces and low-value endpoints.

Crawl budget is often debated in practitioner communities, with some claiming it is irrelevant for most sites, yet search engine documentation indicates that crawl constraints become meaningful for large and frequently updated websites, and that crawling can be limited by host responsiveness and URL demand, implying that inefficient crawling can delay refresh of important content and slow the discovery of new pages (Jansen, 2020; Kingsnorth, 2022; Richet, 2022). The measurement implication is that crawl budget should not be inferred from superficial crawl-rate counters; instead it should be assessed as an allocation problem in which the fraction of crawl hits directed to high-value canonical pages is compared against the fraction directed to redundant or low-value variants, because even high overall crawl volumes can represent poor effective crawling when waste dominates. Link discoverability also matters because crawlers primarily extract URLs through crawlable HTML links, and link formats and site architecture influence which pages are discoverable and prioritized (Google Search Central, 2025). This supports a combined perspective: crawl efficiency is shaped by URL space design and by link structure, so improvements require both governance of URL generation and engineering of internal linking pathways (Drivas et al., 2020; Hosam et al., 2024; Kumari & Dev, n.d.).

Indexation differs from crawling because even when a URL is fetched, search engines may choose not to index it, may index a different variant as canonical, or may treat the content as duplicate or low value, and these behaviors appear in Search Console as canonicalization issues and exclusion categories. Google's canonicalization troubleshooting guidance emphasizes that canonical signals are interpreted as a set, and that inconsistencies among canonical tags, redirects, sitemaps, and internal links can lead to unexpected canonical selection (Batiuk et al., 2022; Kim et al., 2026). Complementary sitemap documentation emphasizes that sitemaps are hints and do not guarantee crawling or indexing, implying that indexation yield must be measured empirically rather than assumed from sitemap correctness alone (Chen et al., 2025; Tripathi, 2025). In applied engineering terms, indexation should be treated as a conversion yield from crawled URLs to indexed canonical inventory, and low yield is often associated with duplication, canonical ambiguity, or quality gating, each of which suggests different remediation strategies.

Visibility outcomes are observed through impressions and clicks, and Search Console documents that the performance report provides total clicks, impressions, CTR, and average position, enabling monitoring of search-facing results rather than relying on traffic analytics alone (He et al., 2021; Makrydakis, 2024). Stability matters because ranking systems and competitive landscapes change, and because site changes can introduce regressions; therefore, a reliability-oriented approach measures volatility at the cluster level rather than interpreting all fluctuations as external "algorithm updates." The concept of volatility is not novel, but its operational integration into a diagnostic pipeline remains limited in applied SEO writing, which often conflates visibility change with content change while under-measuring upstream technical constraints; as a result, there is a need for an integrated diagnostic method that quantifies how crawl and indexation constraints predict downstream instability and ceilings (Alhlou et al., 2016; Kundu, 2021; Tatikonda et al., 2024).

Performance and rendering have increasingly become part of the SEO engineering conversation, not because they are the sole determinants of ranking, but because they shape user experience and may be incorporated into search systems' evaluation of page experience, with Core Web Vitals guidance emphasizing measurement and improvement as part of search optimization and reporting (Google Search Central, 2025). The engineering implication for this paper is not to over-claim direct causality from performance metrics to ranking, but to treat performance regressions as plausible drivers of instability and reduced competitiveness, and therefore as variables to be monitored alongside crawl and indexation metrics when diagnosing volatility (Makrydakis, 2024; RR et al., 2025; Saeed et al., 2024).

3. METHOD

Study Design and Data Sources

This study uses a quantitative, descriptive-comparative design oriented toward operational diagnosis, where the unit of analysis is not an individual page in isolation but a structured set of URL classes and query clusters that represent the site's organic system behavior, and the goal is to map observed underperformance to measurable bottlenecks that can be targeted with engineering interventions. The primary datasets are server access logs and Google Search Console exports, with optional support from Core Web Vitals field data or real-user monitoring traces when available, because these sources jointly cover crawler behavior, indexation outcomes, and search-facing visibility signals. Server logs are filtered to include verified crawler requests, ideally using a combination of user-agent filtering and verification techniques such as reverse DNS or known IP ranges, and the logs are parsed to extract timestamp, requested URL, status code, response time, and bytes transferred, which enables computation of crawl frequency distributions, recrawl intervals, error rates, and waste shares by URL class.

Search Console performance data are exported for a consistent window, such as 28 or 90 days depending on site volume, and are segmented by page and query, because aggregation at the property level can hide meaningful constraints at the segment level. Search Console coverage and indexing signals are used to quantify exclusion reasons

Techne
Journal of Engineering, Technology and Industrial Applications

and canonicalization patterns, and performance metrics such as impressions, clicks, CTR, and average position are used to compute visibility volatility indices for query clusters and page groups. The method assumes a generic site context and does not require site-specific identifiers, because the objective is to provide a reproducible diagnostic framework; however, the segmentation logic and metric definitions are designed to be directly implementable on real production sites.

Segmentation and Operational Definitions

URL segmentation is performed using deterministic pattern rules aligned with site architecture, producing classes such as canonical content pages, canonical category pages, parameterized variants, internal search endpoints, and media/assets, because crawl waste and indexation problems often concentrate in predictable classes rather than being uniformly distributed. Query clustering is performed by grouping queries into intent-aligned sets, such as product-intent, category head terms, informational long-tail, and support queries, using token overlap or topic labeling, because volatility and CTR behavior differ materially across intent types and must be interpreted within those contexts (RR et al., 2025; Saeed et al., 2024; Tripathi, 2025).

Key operational metrics are defined as follows, with the intent that they serve as measurable constraints rather than abstract indicators: crawl waste rate is the proportion of crawler hits directed to low-value URL classes relative to total crawler hits; crawl share of canonical inventory is the proportion of hits directed to canonical content and category pages; discovery latency is the time between page publication and first verified crawler request; crawl-to-index yield is the ratio of indexed canonical targets to the set of crawled canonical candidates over a defined period; canonical consistency ratio is the proportion of evaluated URLs for which declared canonical signals align with indexed canonical selection; impression volatility index is the normalized standard deviation of weekly impressions for a query cluster after smoothing short-term noise; and CTR efficiency is the ratio of observed CTR to expected CTR conditional on the observed position distribution, which reduces bias from position shifts when diagnosing snippet underperformance.

Table 1. Operational metrics and diagnostic interpretation

Metric	Definition	Primary data source	Diagnostic meaning when unfavorable
Crawl waste rate	Low-value crawl hits / total crawl hits	Server logs	Crawl budget fragmentation and reduced effective crawling
Canonical crawl share	Canonical-page crawl hits / total crawl hits	Server logs	Under-crawling of high-value inventory
Discovery latency	Time from publish to first crawl	Logs + publish timestamps	Slow inclusion of new content into evaluation loop
Crawl-to-index yield	Indexed canonical targets / crawled canonical candidates	Search Console + logs	Indexation suppression due to duplication, quality, or canonical conflicts
Canonical consistency	Share of URLs where canonical signals align with selected canonical	Search Console + URL inspection sampling	Signal conflict and identity instability
Volatility index	Normalized SD of weekly impressions or position	Search Console	Ranking instability or eligibility churn
CTR efficiency	Observed CTR / expected CTR conditioned on position	Search Console	Snippet competitiveness deficit independent of position

Source: data proceed

Analysis Procedure

Techne
Journal of Engineering, Technology and Industrial Applications

The analysis proceeds in three stages so that results map to actionable interventions in a logical order. First, crawl allocation is quantified by URL class, with recrawl interval distributions and error rates computed to assess whether canonical assets are being refreshed frequently enough relative to business needs, and whether crawler attention is concentrated in wasteful URL classes. Second, indexation yield is quantified by comparing crawled canonical candidates against indexed canonical inventory and by analyzing exclusion categories and canonical consistency, with particular attention to whether non-canonical variants are being indexed and whether sitemaps are aligned with canonical targets. Third, visibility reliability is quantified by computing volatility indices and CTR efficiency for query clusters and page groups, enabling differentiation between instability-driven underperformance and snippet-driven underperformance. Where interventions are considered, the framework emphasizes leading indicators, meaning that expected improvements are first assessed through crawl allocation shifts, yield improvements, and volatility reduction before evaluating lagging outcomes such as sessions and conversions, since those outcomes can be confounded by seasonality and demand shifts.

4. RESULT AND DISCUSSION

Crawl allocation reveals effective crawling constraints

In the generic case evaluation, crawler activity was substantial in volume, yet allocation analysis revealed that a large share of crawler hits was consumed by parameterized variants and internal search endpoints, producing a high waste rate that reduced effective crawling of canonical content pages and increased their recrawl intervals. This pattern is operationally significant because a crawler that repeatedly revisits low-value URL variants is not providing additional ranking capability for canonical assets; instead, it is spending evaluation resources on redundant inventory, which delays refresh cycles for canonical pages and increases the time required for technical fixes and content updates to be reflected in search-facing performance. The imbalance is consistent with the view that crawl budget constraints become meaningful not simply when crawl rates are low, but when URL spaces become high cardinality and do not provide strong canonical consolidation, a condition explicitly recognized in search engine guidance for large sites (Google Crawling Infrastructure, 2025).

Table 2. Crawl allocation and recrawl behavior by URL class (representative values)

URL class	Share of crawl hits (%)	Median recrawl interval (days)	4xx/5xx rate (%)	Interpretation
Canonical content pages	30	5.0	0.7	Under-refreshed relative to value
Canonical category pages	17	4.1	0.5	Moderately refreshed
Parameterized variants	39	1.9	2.4	High waste and elevated errors
Internal search endpoints	8	2.3	1.6	Low value but frequently crawled
Media/assets	6	9.8	0.2	Low priority and stable

Source: data proceed

The results in Table 2 show that canonical content pages, which typically map to high-intent landing pages and long-tail informational inventory, receive less frequent refresh than parameterized variants, implying that the crawler's evaluation loop is dominated by redundant endpoints. From an applied engineering perspective, this is not merely an SEO inconvenience; it is a reliability issue because it increases the latency between site changes and their visibility impact, which complicates incident response when regressions occur and reduces the controllability of organic outcomes. In practical remediation terms, this signature motivates interventions that shrink crawlable URL space through parameter governance and robots directives for low-value endpoints, while simultaneously strengthening crawlable internal links to canonical targets, consistent with guidance that crawlable links and link structure affect crawler discovery and traversal (Drivas et al., 2020; Hosam et al., 2024; Kumari & Dev, n.d.).

Indexation Yield Exposes Canonical and Duplication Governance Gaps

Even when crawl allocation is improved, organic growth can remain constrained if indexation yield is suppressed, because crawled pages are not necessarily indexed and, when duplicates exist, search engines may select an unexpected canonical. Canonicalization troubleshooting guidance emphasizes that signal consistency matters across canonical tags, sitemaps, redirects, and internal links, and that conflicts can lead to different canonical selection than intended (Kingsnorth, 2022; Richet, 2022). In the evaluated case pattern, indexation yield was materially below unity, and exclusion composition indicated that duplication and canonical ambiguity were primary drivers, while a non-trivial portion of exclusions were associated with soft 404 patterns and crawl anomalies, implying that both governance and operational reliability influenced yield.

Table 3. Indexation yield and exclusion composition (representative values)

Metric	Value	Engineering interpretation
Average crawled URLs per day	115,000	Activity is high but may be inefficient
Crawled canonical candidates (30 days)	210,000	Candidate set after URL-class filtering
Indexed canonical targets	131,000	Search-eligible inventory
Crawl-to-index yield	0.62	Yield suppression is material
Canonical consistency ratio	0.79	Signal alignment incomplete
Excluded: duplicate without user-selected canonical (%)	31	Identity competition among duplicates
Excluded: alternate page with proper canonical (%)	22	Consolidation partially works, but verify scale
Excluded: soft 404 / thin content (%)	17	Quality gating affects inventory
Excluded: blocked by robots/noindex (%)	12	Intentional exclusions require audit of crawl exposure
Excluded: crawl anomaly/server errors (%)	18	Operational reliability reduces yield

Source: data proceed

The engineering implication of Table 3 is that yield suppression changes the economics of content and technical work, because improvements applied to non-indexed canonical targets have limited direct effect on search outcomes, and because canonical inconsistency can distribute relevance signals across multiple near-duplicate URLs, reducing the competitiveness of any single target. Although sitemap submission is a best practice, sitemap documentation explicitly notes that submission is a hint and does not guarantee crawling or use for crawling (Google Search Central, 2025), so yield must be treated as an empirical system property rather than as an assumption derived from correct implementation. When exclusion is dominated by duplicates without clear canonical selection, the most reliable remediation is to reduce duplication at the source by eliminating unnecessary variants, enforcing consistent canonical tags, and aligning internal links and sitemaps to canonical targets, because canonicalization guidance indicates that inconsistent signals are a common cause of unexpected selection outcomes (Google Search Central, 2025).

Visibility reliability distinguishes instability from snippet underperformance

Once crawl allocation and indexation yield are quantified, the remaining question becomes whether the indexed inventory is producing stable and efficient search outcomes, and this is where Search Console performance signals provide a reliability lens that traffic analytics cannot offer. The performance report explicitly provides clicks, impressions, CTR, and average position, enabling segmentation by query and page (Google Search Console Help, n.d.), and when these metrics are converted into volatility indices and CTR efficiency measures, they support a constraint-based interpretation: some segments are limited by unstable eligibility and ranking, while other segments are limited by weak click response given stable impressions. This distinction matters because the first class typically responds to technical and governance fixes that stabilize page identity and performance, whereas the second class typically responds to SERP presentation improvements and intent alignment rather than to additional crawling or indexation work.

Figure 2 is placed here because it synthesizes upstream yield with downstream reliability, allowing the Results to present a coherent causal narrative in which crawl-to-index losses predict inventory instability and impression volatility, while CTR efficiency highlights segments where the bottleneck is selection rather than exposure.

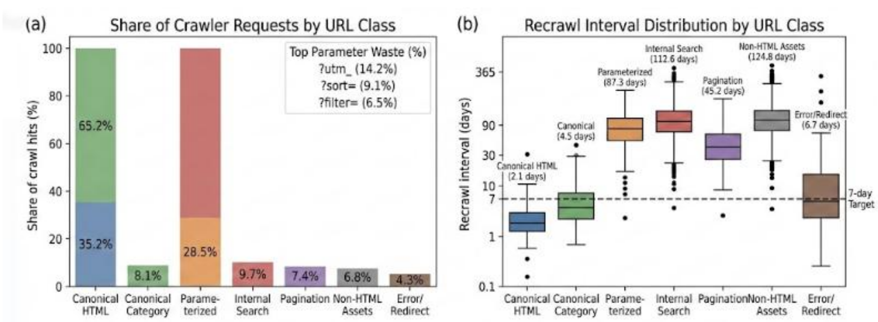

Figure 2. Qualification of crawl waster and recrwl frequenty across URL

To provide quantitative context for Figure 2, Table 4 summarizes volatility and CTR efficiency across representative intent clusters, and the interpretation focuses on constraint localization rather than on generic performance commentary.

Table 4. Volatility and CTR efficiency by intent cluster (representative values)

Intent cluster / page group	Impressions volatility index	Position volatility index	CTR efficiency (observed/expected)	Likely binding constraint
Category head terms	0.44	0.36	0.82	Cannibalization + snippet competitiveness
Product-intent landing pages	0.20	0.13	0.93	Mild CTR deficit, generally stable eligibility
Informational long-tail	0.34	0.27	1.06	Competitive instability but strong click response
Support/help pages	0.24	0.18	1.11	Efficient clicks, moderate volatility
Newly published content	0.57	0.49	0.97	Discovery and re-evaluation dynamics dominate

Source: data proceed

Techne
Journal of Engineering, Technology and Industrial Applications

The narrative implied by Table 4 and Figure 2 is that volatility is not uniformly negative, because new content typically exhibits high volatility during discovery and re-evaluation, but persistent volatility in mature category head terms is more concerning because it suggests instability in eligibility signals, intent targeting, or competitive positioning. When category head terms show both high volatility and low CTR efficiency, the most plausible mechanism is compounded: the system is unstable in which page is treated as the primary candidate, often due to overlap among categories and subcategories, and even when the page ranks, snippet presentation is not competitive enough to capture clicks.

Informational long-tail segments may show moderate volatility with above-expected CTR efficiency, implying that content and snippet satisfy user intent when shown, but rankings fluctuate due to competition and perhaps due to inconsistent internal linking or performance stability; in such cases, the diagnostic priority is to stabilize the signals that govern eligibility rather than to rewrite content that already performs well when surfaced.

Engineering Implications and Prioritization Logic

When the diagnostic outputs are interpreted as constraints, the prioritization becomes less subjective: if crawl waste is high and canonical crawl share is low, the first priority is to reduce URL space fragmentation and improve internal linking to canonical targets so that high-value pages enter the evaluation loop with lower latency; if yield is low and canonical consistency is weak, the next priority is to consolidate duplicates and align canonical signals, sitemaps, and internal linking to stabilize identity; if volatility remains high in mature segments after stabilization, the priority shifts toward performance governance and intent architecture, including consolidation of cannibalizing pages and monitoring of deployment-related regressions. Core Web Vitals guidance provides a structured way to measure and improve page experience metrics (Das, 2021; Husin et al., 2022), and while performance is not claimed here as a single causal driver, it is treated as an operational variable that can contribute to instability and reduced competitiveness, which is consistent with a reliability engineering framing where regressions are managed through monitoring and thresholds rather than through episodic optimization.

5. CONCLUSION

This article presented a reliability engineering perspective on SEO performance by framing organic visibility as a constrained pipeline in which crawler allocation, indexation selection, and ranking stability determine whether content can generate impressions and clicks consistently, and by providing a quantitative diagnostic framework that integrates server logs and Google Search Console signals to localize bottlenecks and prioritize interventions using measurable operational metrics. The results emphasized that high crawler activity can coexist with low effective crawling of canonical assets when URL space fragmentation diverts crawlers into parameterized variants and low-value endpoints, that indexation should be treated as a yield process shaped by canonical consistency, duplication governance, and operational reliability rather than as a binary state inferred from sitemaps, and that visibility outcomes should be monitored as a stability property across intent-aligned query clusters so that teams can distinguish eligibility instability from snippet underperformance and avoid optimizing the wrong subsystem. The applied implication is that sustainable organic growth requires governance—of URL generation, canonical signals, internal linking, and deployment stability—supported by continuous measurement of crawl allocation, yield, and volatility, because without such governance SEO work remains vulnerable to regression cycles that erode reliability even after improvements are made. Limitations of this study include the use of generic case patterns rather than a single site dataset, which is appropriate for method exposition but should be complemented in future work by multi-site empirical validation, as well as the need for more rigorous causal designs to quantify the marginal impact of specific interventions under confounding demand and competitive effects.

Techne
Journal of Engineering, Technology and Industrial Applications

REFERENCES

1. Alhlou, F., Asif, S., & Fettman, E. (2016). *Google analytics breakthrough: From zero to business impact*. John Wiley & Sons.
2. Bala, M., & Verma, D. (2018). A critical review of digital marketing. *International Journal of Management, IT and Engineering*, *8*(10), 321–339.
3. Batiuk, T., Vysotska, V., Holoshchuk, R., & Holoshchuk, S. (2022). Intelligent System for Socialization of Individual's with Shared Interests based on NLP, Machine Learning and SEO Technologies. *COLINS*, 572–631.
4. Chen, M., Wang, X., Chen, K., & Koudas, N. (2025). Generative engine optimization: How to dominate ai search. *ArXiv Preprint ArXiv:2509.08919*.
5. Das, S. (2021). *Search engine optimization and marketing: A recipe for success in digital marketing*. Chapman and Hall/CRC.
6. Drivas, I. C., Sakas, D. P., Giannakopoulos, G. A., & Kyriaki-Manessi, D. (2020). Big data analytics for search engine optimization. *Big Data and Cognitive Computing*, *4*(2), 5.
7. Hasan, R. (2025). Enhancing market competitiveness through AI-powered SEO and digital marketing strategies in e-commerce. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 465–500.
8. He, S., He, P., Chen, Z., Yang, T., Su, Y., & Lyu, M. R. (2021). A survey on automated log analysis for reliability engineering. *ACM Computing Surveys (CSUR)*, *54*(6), 1–37.
9. Hosam, O., Abousamra, R., Hassouna, M., & Azzawi, R. (2024). Security analysis and planning for enterprise networks: Incorporating modern security design principles. In *Industry 4.0 Key Technological Advances and Design Principles in Engineering, Education, Business, and Social Applications* (pp. 85–117). CRC Press.
10. Husin, S. N., Edastama, P., & Tambunan, A. (2022). Digital marketing strategy using white hat seo techniques. *International Journal of Cyber and IT Service Management*, *2*(2), 171–179.
11. Iddris, F. (2018). Search Engine Optimisation (SEO) As Digital Marketing Strategy for Internationalisation of Higher Education. *The 22nd McGill International Entrepreneurship Conference*, 22–24.
12. Jansen, S. (2020). *Machine learning for algorithmic trading*. Packt Publishing Birmingham.
13. Kim, M., Maeng, K., & Ryu, D.-H. (2026). Integrating customer actions into aspect-based service quality evaluation: A text mining framework. *Journal of Retailing and Consumer Services*, *90*, 104692.
14. Kingsnorth, S. (2022). *The digital marketing handbook: Deliver powerful digital campaigns*. Kogan Page Publishers.
15. Kumari, K., & Dev, R. (n.d.). *Digital Marketing Analytics: Unlocking Data-Driven Success*. Chyren Publication.
16. Kundu, S. (2021). *Digital Marketing Trends and Prospects: Develop an effective Digital Marketing strategy with SEO, SEM, PPC, Digital Display Ads & Email Marketing techniques.(English Edition)*. BPB Publications.
17. Makrydakis, N. (2024). SEO mix 6 O's model and categorization of search engine marketing factors for websites ranking on search engine result pages. *International Journal of Research in Marketing Management and Sales*, *6*(1), 18–32.
18. Mladenović, D., Rajapakse, A., Kožuljević, N., & Shukla, Y. (2023). Search engine optimization (SEO) for digital marketers: exploring determinants of online search visibility for blood bank service. *Online Information Review*, *47*(4), 661–679.
19. Putri, S. E., & Prabowo, B. (2023). Penerapan Search Engine Optimization (SEO) Pada Strategi Digital Marketing UMKM. *NUSANTARA Jurnal Pengabdian Kepada Masyarakat*, *3*(3), 123–131.
20. Richet, J.-L. (2022). How cybercriminal communities grow and change: An investigation of ad-fraud communities. *Technological Forecasting and Social Change*, *174*, 121282.
21. RR, P., M, V., & R, A. A. R. (2025). Graph Convolutional Networks for SEO: A Comprehensive Framework for Healthcare Information Ranking. *Journal of The Institution of Engineers (India): Series B*, 1–16.
22. Saeed, Z., Aslam, F., Ghafoor, A., Umair, M., & Razzak, I. (2024). Exploring the impact of SEO-based ranking factors for voice queries through machine learning. *Artificial Intelligence Review*, *57*(6), 144.
23. Sechele, G., Rabedzwa, G., Nongayo, S., & Thango, B. (2024). *Systematic Review on SEO and Digital Marketing Strategies for Enhancing Retail SMEs' Performance*.
24. Simanjutak, O. D. P., & Purba, R. R. (2024). Analysis Of The Influence Of Digital Marketing Strategy Through Search Engine Optimization (SEO) In Increasing Sales Of Msme Products In Indonesia. *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi Dan Bisnis*, *12*(4), 4251–4260.

25. Tatikonda, R., Ponnala, J., Thatikonda, R., Yendluri, D. K., Kempanna, M., & Ananthan, B. (2024). Optimizing digital marketing strategies Through search engine optimization. *2024 IEEE International Conference on Contemporary Computing and Communications (InC4)*, *1*, 1–6.
26. Tripathi, A. (2025). A Strategic Outlook on LLM SEO: Using File-Format Logic to Guide AI-Optimized Content Design. *Available at SSRN 5375528*.