

Click Happens

On-Page SEO Analysis

Sample Deliverable · client.com

Prepared by
Dominik Liebezeit

Scope
**Technical & On-Page
SEO**

Executive Summary

The website is technically functional, but significantly underperforms due to critical issues in indexation, page speed, and content. Estimated loss: 50–70% of potential organic traffic.

Top Issues

Issue	Business Impact	Reference
70% of pages not indexed by Google (13,000+ URLs)	Massive traffic loss — products Google doesn't know don't sell	p. X — Indexation
Slow load times (3.5s–4.2s)	Lower rankings + higher bounce rate = fewer conversions	p. X — Page Speed
Categories and products without quality content	Missing long-tail traffic, invisible for informational queries	p. X — Content

Business Impact

- Losing the majority of potential customers from Google
- Competitors with better SEO are pulling ahead
- Investment in product catalogue isn't paying off (products are invisible)
- Slow site = more abandoned carts

Recommended Approach & Investment

Optimisation in 3 phases, total estimated effort: 140–190 hours (developer, SEO specialist, copywriter):

Phase	Timeline	Effort
Phase 1: Quick wins (speed, schema, titles)	Month 1	40–60h
Phase 2: Content optimisation (copy, descriptions)	Months 2–3	60–80h
Phase 3: Technical optimisation (filters, server)	Months 3–4	40–50h

Expected Outcome

6-MONTH PROJECTION

+55–80% organic traffic growth

- Better rankings (Core Web Vitals as ranking factor)
- Higher click-through rates (+25–35% via rich snippets)
- Higher conversion rate (+10–15% from faster site speed)
- Long-term stable organic search performance

Indexation

Indexation is the process by which search engines (Google, Bing) crawl a website, analyse its content, and store information in their database. Only indexed pages can appear in search results.

KEY IMPACTS

Visibility: Non-indexed pages = zero organic traffic. Business impact: Every non-indexed product page represents lost potential conversions. Competitive disadvantage: A poorly indexed site surrenders ground to competitors.

Sitemap.xml

Sitemap.xml is a structured XML file containing a list of all important URLs on the site. It acts as a "map" for search engines, helping them discover the pages that matter.

- More efficient crawling — search engines discover all important pages faster
- Faster indexation — new content can enter the index more quickly
- Monitoring — provides data on which pages have actually been indexed

FINDING

More than 65% of pages in the sitemap are not indexed by Google. In absolute terms, that is over 13,000 pages.

Only pages we want search engines to know about should be in sitemap.xml. When a significant share is rejected, it either means the sitemap contains unimportant pages that shouldn't be there, or — as is the case here — search engines are refusing to add them to their database.

Google vs. Seznam comparison: Seznam (the Czech search engine) indexes more than 97% of the same pages. This discrepancy typically indicates that pages are not sufficiently unique and Google is flagging them as duplicates. Seznam is significantly less strict and will index almost any URL it can access that it hasn't seen before.

To precisely identify why Google is refusing to index these pages, access to Google Search Console is required. The root cause is most likely one or a combination of the following:

1. Page title is not sufficiently unique
2. Pages are not sufficiently internally linked

The first step is to establish what type of pages these are, how they were created, and whether we want them indexed at all. A list of affected URLs is attached as: client_non_indexed_urls.xlsx

- If these pages should NOT be indexed — canonicalise them and remove from sitemap
- If they SHOULD be indexed — diagnose uniqueness and internal link issues

Robots.txt

Robots.txt is a text file at the root of the website (client.com/robots.txt) that defines rules for crawler access to different parts of the site. It acts as a "gatekeeper" controlling which sections search engine bots can access.

- Crawl control — direct which sections of the site are crawled
- Crawl budget optimisation — prevent bots wasting time on irrelevant pages (duplicates)
- Protection of sensitive data — prevent indexation of internal tools and staging pages
- Server performance — reduce server load by blocking unnecessary crawls

FINDING

client.com currently does not block crawling of sort/filter parameter URLs or internal search result pages via robots.txt. This causes crawl budget waste, duplicate content, and index bloat.

Examples of problematic URLs being crawled:

- /search?q=some+query
- /search?term=product&page=2&filter=price
- /catalogsearch/result/?q=xyz&p=5

Recommended robots.txt configuration:

```
User-agent: *
Disallow: /*/collections/*sort_by*
Disallow: /collections/**
Disallow: /collections/*%2B*
Disallow: /search*
Sitemap: https://client.com/sitemap.xml
```

Meta Robots — Blocking Indexation

The meta robots tag is an HTML element in the <head> section that gives direct instructions to search engines on how to handle a page. Unlike robots.txt (which only blocks crawling), meta robots directly blocks indexation.

Key distinction:

- Robots.txt: "Don't come here" (prevents crawling)
- Meta robots: "You can come, but don't add this page to the index"

Meta robots in code:

```
<meta name="robots" content="noindex, nofollow">
```

FINDING

Meta robots is correctly implemented on most page types. One area requires attention: internal search result pages are currently being indexed.

Example affected URLs:

- <https://client.com/search?q=television>
- <https://client.com/search?term=laptop&page=2>

Why this is a problem:

3. Duplicate content at scale — every search query creates a new URL with similar product listings
4. Thin content — search pages rarely contain unique content valuable to users arriving from Google
5. Crawl budget waste — Google spends time on thousands of search query combinations instead of new products
6. Unintended SERP appearance — users see "Search results" pages instead of specific products

Recommended fix: Implement noindex, follow on all internal search result pages:

```
<meta name="robots" content="noindex, follow">
```

HTTP Status Codes & Page Availability

HTTP status codes are numeric responses from the server informing browsers and search engines about the status of a requested page. Correct configuration is critical for SEO — it affects whether and how pages are indexed, how authority flows between URLs, and user experience.

Status Code Reference

Code	Meaning	SEO Impact
200 OK	Page is available, server returned it successfully	Positive — all accessible pages should return 200
301 Moved Permanently	Page has permanently moved to a new URL	Passes link equity — use for all permanent URL changes
404 Not Found	Requested page does not exist	Neutral if intentional, harmful if pages have traffic or links

Findings on client.com

FINDING

23 URLs returning 404 still generate organic traffic according to Google Search Console. 47 internal links point to non-existent pages. 12 URLs in sitemap.xml return 404.

404 Pages with Organic Traffic

URL	Monthly Visits	Status
/products/lg-oled-tv-2022/	340	404
/category/laptops-under-20000/	180	404
/blog/mobile-buying-guide/	95	404

Recommendation: Restore the page or implement a 301 redirect to the nearest equivalent.

Internal Links Pointing to 404

- Homepage → "Deals of the week" → 404 (expired promotion)
 - Category → Related product → 404 (discontinued model)
 - Blog → Internal reference → 404 (deleted article)
 - Footer → "About us" → 404 (old version of page)
7. Fix URL — update links to correct target pages
 8. Remove links — for permanently invalid targets (promotions, campaigns)
 9. Implement 301 — for moved pages
 10. Regular audit — monthly check of internal broken links

404 URLs in sitemap.xml

Sitemap.xml contains 12 URLs returning 404. Action steps:

11. Remove all 404 URLs from sitemap
12. Implement automated sitemap validation (sitemap should only contain 200 OK URLs)
13. Re-submit sitemap in Google Search Console

Canonicalisation

The canonical tag (`rel="canonical"`) is an HTML element that designates the preferred version of a page when multiple URLs exist with identical or very similar content. It tells search engines: "This is the original — index this and assign it all authority."

```
<link rel="canonical" href="https://client.com/preferred-url/" />
```

Correct canonicalisation is important because it:

- Resolves duplicate content without requiring 301 redirects
- Consolidates link equity from different URL versions onto one preferred version
- Prevents competing pages (multiple versions of the same page competing in search)
- Allows different URL versions for users while presenting a single version to search engines

STATUS: PASS

Canonicalisation is correctly implemented on the site. Self-referencing canonical tags are in place across all preferred page types (products, categories, articles), providing protection against duplicate content from UTM parameters, session IDs, and other tracking codes.

Product Filters & Category Facets

Filters are a key part of e-commerce UX, allowing users to find exactly what they need. For SEO they can also create valuable landing pages. However, each filter combination creates a new URL, which can lead to massive duplicate content and crawl budget waste.

The maths of filter combinations:

EXAMPLE

5 brands × 4 price ranges × 6 attributes × 2 availability states = 240 possible URL combinations for a single category page

Findings

FINDING

Google Search Console shows approximately 1,458,800 indexed filter URLs across various categories on client.com.

Example indexed filter URLs:

- <https://client.com/category/televisions/?brand=samsung&size=55>
- <https://client.com/category/televisions/?brand=samsung&price=20000-30000>
- <https://client.com/category/televisions/?brand=lg&size=65&price=30000-50000>

Identified problems:

14. Massive index bloat — over 1.4M indexed filter pages dilute crawl budget
15. Link equity fragmentation — external links pointing to different filter variants split authority
16. Crawl budget waste — Google spends resources on filter permutations instead of new products

A detailed filter analysis is required to map all filter combinations and design a new filter logic. This should be preceded by comprehensive keyword research to determine which filter combinations represent genuine search demand.

Page Speed

Page speed is a confirmed ranking factor and directly affects user experience. Google measures performance using Core Web Vitals — a set of metrics focused on real user experience.

Core Web Vitals Explained

Metric	What It Measures	Good Threshold
LCP — Largest Contentful Paint	How quickly the main content loads (hero image, main heading)	< 2.5s
INP — Interaction to Next Paint	Responsiveness to user interaction	< 100ms
CLS — Cumulative Layout Shift	Visual stability during page load	< 0.1

Current Performance on client.com

Homepage — <https://client.com/>

Metric	Value	Status	Target
LCP	3.5s	⚠ Needs improvement	< 2.5s
CLS	0.5	✗ Poor	< 0.1
INP	85ms	✓ Good	< 100ms

Overall scores: Mobile 52/100 (Poor) · Desktop 68/100 (Average)

Product Page — <https://client.com/products/samsung-qled-tv/>

Metric	Value	Status	Target
LCP	4.2s	✗ Poor	< 2.5s
CLS	0.65	✗ Poor	< 0.1
INP	110ms	⚠ Needs improvement	< 100ms

Overall scores: Mobile 45/100 (Poor) · Desktop 61/100 (Average)

Root Cause Analysis

Problem 1 — LCP 3.5s (Homepage) and 4.2s (Product)

LCP measures how quickly the largest visible element loads (typically the hero banner or main product image).

17. Unoptimised images

- Hero banner: 2.8 MB (should be max 200–300 KB)
 - Product photos: average 800 KB (should be 80–150 KB)
 - No modern WebP/AVIF format in use
 - No responsive images (same size served to mobile and desktop)
 - No lazy loading implemented
18. Render-blocking resources (CSS/JS)
- Browser must download and process 1.2 MB CSS/JS before rendering the page
19. Slow server response time
- Time to First Byte (TTFB): 850ms (should be < 200ms)
 - Possible causes: slow hosting, no server-side caching, unoptimised database queries, no CDN
20. Missing critical optimisations
- No preload for critical fonts
 - No preconnect for external domains
 - No critical CSS inlining

Problem 2 — CLS 0.5 (Homepage) and 0.65 (Product)

CLS measures visual stability — how much content shifts during page load. Current scores are 5–6.5× worse than the acceptable threshold.

21. Images without defined dimensions
- ```

```
22. Dynamically loaded content without reserved space
- Ads/banners loaded asynchronously collapse to 0px height until loaded, then cause a layout shift
  - Fix: reserve space with min-height before content loads
23. Web fonts causing text reflow
- Text renders in fallback font first, then jumps when the web font loads
  - Fix: font-display: swap + preload critical fonts

## Content

Quality content is the foundation of successful SEO. Even with perfect technical implementation, a site won't rank without relevant, optimised content. This section analyses the content across key page types on client.com.

Pages audited: Homepage, Category pages (televisions, laptops, mobiles), Product pages, Blog articles, About / Contact pages.

### Identified Issues

#### Issue 1 — Thin Content on Category Pages

The majority of category pages contain only a product listing with no editorial content.

- No context for search engines (what is the page actually about?)
- Low relevance for informational queries
- Missed opportunity for long-tail keyword traffic
- No value to users beyond the product grid

#### Issue 2 — Duplicate Product Descriptions

85% of product descriptions are taken directly from manufacturers — identical to competitor sites.

- Google cannot determine which site to prioritise
- No unique value for users
- Lower rankings due to duplicate content
- No opportunity for brand differentiation

#### Issue 3 — Blog Without a Strategy

| Metric                 | Current State | Competitor Benchmark |
|------------------------|---------------|----------------------|
| Last article published | 8 months ago  | 2–4 articles/month   |
| Total articles         | 12            | 50+                  |
| Average length         | 350 words     | 1,200+ words         |
| Topic strategy         | Ad hoc        | Topical clusters     |

## Page Titles

The title tag is the most important on-page SEO element. It appears as the blue clickable link in search results and directly impacts:

- Ranking — Google uses titles to understand the topic of a page
- CTR — a well-written title increases clicks by 20–40%
- Relevance — must include keywords that people actually search for

Optimal length: 50–60 characters. Recommended structure:

[Primary keyword] | [Context] | [Brand]

### FINDING

Audit of 50 randomly selected pages found titles are not optimised. Keywords are not used effectively and some titles are duplicated. The problem is most prevalent on category pages and blog articles.

## Meta Descriptions

The meta description appears beneath the title in search results. Although not a direct ranking factor, it has a significant impact on CTR (+20–40% for well-written descriptions) and communicates relevance to users.

Optimal length: 150–160 characters.

## Audit Findings

| Page Type  | Total | Missing   | Duplicate | Too Short | OK       |
|------------|-------|-----------|-----------|-----------|----------|
| Products   | 450   | 180 (40%) | 120 (27%) | 80 (18%)  | 70 (15%) |
| Categories | 18    | 12 (67%)  | 4 (22%)   | 2 (11%)   | 0 (0%)   |
| Blog       | 12    | 8 (67%)   | 2 (17%)   | —         | 2 (16%)  |

65% of pages have a meta description problem.

Recommended approach:

- Categories and key products — write manually
- Less important products and blog articles — generate with AI, review before publishing
- All remaining products — let Google auto-generate from page content

# Heading Structure

Correct heading structure helps Google understand the hierarchy of content and increases the chance of featured snippets. Headings carry more SEO weight and improve readability for users.

Basic rules:

- One H1 per page
- Logical hierarchy: H1 → H2 → H3 (no skipping levels)
- Keywords in headings

## Identified Issue — Category Pages

| Current State                  | Recommended Structure                    |
|--------------------------------|------------------------------------------|
| H1: Televisions                | H1: Televisions — Wide Selection from £X |
| H3: Samsung QE55Q80C (product) | H2: How to Choose a Television           |
| H3: LG OLED55C3 (product)      | H3: Screen size considerations           |
| H3: Sony XR-55A80L (product)   | H3: OLED vs QLED panel types             |
| 48× H3 product headings        | H2: Our Television Range                 |
| No H2 headings                 | H2: Frequently Asked Questions           |
| No content structure           | H3: What size TV is right for my room?   |

## Internal Linking

Internal links (links between pages on your own site) are critical for crawling (helping bots discover all pages), authority distribution (strong pages passing value to weaker ones), user navigation, and providing context to Google via anchor text.

### Findings

#### Issue 1 — Orphan Pages

##### FINDING

23 pages have zero internal links pointing to them. They are accessible only via sitemap.xml or external links. Google struggles to find and index them.

#### Issue 2 — Weak Category Linking Structure

- Homepage links to each category only once
- Category pages don't link to related categories
- Product pages don't link to accessories
- Blog articles don't link to products or categories

Example: The article "How to Choose a Television" contains NO links to the television category or any specific products.

#### Issue 3 — Generic Anchor Texts

180+ links use non-descriptive anchor text:

| Anchor Text        | Occurrences | Problem                                               |
|--------------------|-------------|-------------------------------------------------------|
| "More information" | 85×         | Google cannot determine what the target page is about |
| "View details"     | 62×         | No keyword signal in anchor text                      |
| "Click here"       | 33×         | Wasted opportunity to pass topical context            |

Recommendation: Replace generic anchors with descriptive keyword-rich text (e.g. "LG OLED televisions" instead of "More products here").

#### Issue 4 — Insufficient Product Cross-Linking

Product pages currently only contain: breadcrumbs and 4 randomly selected related products.

Missing:

- Links to accessories
- Links to alternatives (cheaper/more expensive variants)
- Links to comparison guides
- Links to relevant blog articles

## Structured Data (Schema.org)

Structured data (schema markup) is code added to HTML that helps search engines better understand page content. It enables rich snippets — enhanced search results showing star ratings, prices, availability, and other information directly in the SERP.

### AI SEARCH NOTE

With users increasingly turning to AI tools (ChatGPT, Gemini, Claude, Perplexity) instead of traditional search, structured data is more important than ever — it helps these tools understand and cite your content accurately.

Key benefits:

- AI tools — helps your site appear as a cited source in AI-generated answers
- Rich snippets — stars, prices, availability directly in SERP
- Higher CTR — enriched results attract more clicks (+20–30%)
- Better Google understanding — precise content interpretation
- Competitive advantage — more prominent SERP display than competitors

### Current State on client.com

| Schema Type          | Status                        | Priority   |
|----------------------|-------------------------------|------------|
| Organization (basic) | ✅ Implemented (homepage)      | —          |
| Product schema       | ❌ Missing (0 of 450 products) | Priority 1 |
| AggregateRating      | ❌ Missing                     | Priority 1 |
| BreadcrumbList       | ❌ Missing                     | Priority 2 |
| FAQPage              | ❌ Missing                     | Priority 3 |
| LocalBusiness        | ❌ Missing                     | Priority 4 |

### Implementation Priorities

#### Priority 1 — Product Schema

Add to all product pages. Minimum required fields:

- @type: Product
- name, image, description, sku, brand
- offers: price, priceCurrency, availability, url
- aggregateRating: ratingValue, reviewCount (if ratings exist)

#### Priority 2 — BreadcrumbList

Add to all product and category pages. Structure: Homepage → Category → Subcategory → Product. Each level requires position and url properties.



### Priority 3 — FAQPage

Add to blog articles with FAQ sections, category pages with FAQ content, and products with Q&A sections.

### Priority 4 — LocalBusiness

Add to homepage and contact page. Required fields: name, address, telephone, openingHours, priceRange.

## Implementation Guidelines

- Format: JSON-LD (recommended by Google). Place in <head> or end of <body>
- Products: generate schema dynamically from product data (price, availability, name, SKU)
- Categories: BreadcrumbList + FAQPage (if FAQ content exists)
- Blog: Article + FAQPage (if FAQ section present)
- Homepage: Organization + LocalBusiness

Validation tools:

- Google Rich Results Test: [search.google.com/test/rich-results](https://search.google.com/test/rich-results)
- Schema Markup Validator: [validator.schema.org](https://validator.schema.org)
- Google Search Console → Enhancements

After implementation: test 10 random products and all page types, then request indexing via GSC for faster processing.

## Questions about this analysis?

[clickhappens.cz](https://clickhappens.cz)

Dominik Liebezeit · Senior SEO Specialist