GOBIYA TECHNICAL SERIES

The Architecture Handbook

Technical SEO and search architecture — the foundation everything else is built on

Written by **Steve Martin**, Founder of Gobiya

CONTENTS

Table of Contents

CHAPTER 01

Why Architecture Sets Your Ceiling

Content gets the attention. Architecture sets the limit on what that content can ever achieve.

You can write the best page in your industry, and if a search engine cannot crawl it, cannot render it, or cannot work out how it relates to the rest of your site, that page will underperform a mediocre one on a well-built competitor's domain. This is the single most common reason good content fails.

Architecture is the ceiling. Content is how close you get to it.

The good news is that technical problems are the most fixable category in search. They are concrete, they are checkable, and they usually stay fixed. Unlike rankings, which move for reasons outside your control, a rendering bug you repair today is repaired tomorrow.

What Bad Architecture Actually Costs

Technical problems rarely announce themselves. Nothing breaks, no error appears, and the site looks fine to everyone who works on it. The cost shows up somewhere else entirely.

1. **Content that never had a chance.** Every page published onto a site machines cannot properly read is money spent on something nobody will find. The writing was fine. It was never the problem.
2. **A ceiling nobody can explain.** Rankings improve to a point and then stop, and no amount of additional content moves them, because the constraint was never content.
3. **Silent decay.** A redesign strips your structured data, a plugin update slows every page, a URL changes without a redirect. Each one costs you quietly for months before anyone notices.
4. **Rework instead of progress.** Teams end up rebuilding the same pages two and three times, because the foundation was never fixed and each attempt hits the same wall.

The pattern is consistent: architecture problems are cheap to prevent, moderate to fix, and expensive to ignore. Everything in this handbook is aimed at the first two.

CHAPTER 02

Crawlability: Getting Found at All

Before a page can rank, be cited, or convert anyone, a machine has to be able to reach it. Crawlability is that first gate, and more sites fail it than their owners realize.

Four things decide whether a bot can reach a page:

1. **robots.txt** — a single stray disallow line can hide an entire section of your site. This file is checked before anything else.
2. **Internal links** — a page with no links pointing at it is effectively invisible, no matter how good it is.
3. **Sitemap** — your XML sitemap should list every page you want indexed, and nothing you do not.
4. **Status codes** — a page returning an error or an unexpected redirect will not be indexed, however fine it looks in a browser.

Check these in order. Fixing content problems while a robots.txt rule is blocking the section is wasted effort.

The Orphan Page Problem

An orphan page is one that exists, loads fine, and has no internal links pointing to it. It usually happens by accident: a page gets built for a campaign, the campaign ends, the navigation changes, and the page is left stranded.

Search engines find pages mainly by following links. A page nothing links to has to be discovered some other way — usually your sitemap — and even when it is found, it inherits none of the authority your other pages have accumulated.

How to find them: compare the list of pages in your sitemap against the pages reachable by clicking through your own site. Anything in the first list but not the second is an orphan.

How to fix them: either link to the page from somewhere relevant, or remove it. A page worth keeping is worth linking to. A page not worth linking to is usually not worth keeping.

Index Bloat

The opposite problem is just as damaging: thousands of low-value pages getting indexed and diluting the site.

Common sources are filtered and sorted URL variants, internal search results, tag and category archives with one item in them, printer-friendly duplicates, and staging pages that were never blocked.

SYMPTOM	LIKELY CAUSE	FIX
Far more indexed pages than real ones	URL parameters	Canonical tags, parameter rules
Near-identical pages ranking against each other	Thin archives	Consolidate or noindex
Staging content in results	Unprotected environment	Password protection, not just noindex
Search result pages indexed	Internal search crawlable	Disallow in robots.txt

The goal is not the largest possible index. It is an index that contains your good pages and nothing else.

CHAPTER 03

Rendering and the JavaScript Trap

This is the most expensive technical mistake we find, and the hardest for a non-specialist to spot — because the page looks perfect in your browser.

Modern sites often build their content with JavaScript after the initial page loads. Your browser waits patiently and runs it. Some bots do too. Many do not, and several of the AI crawlers that matter most for citations do not execute JavaScript at all.

If your content only exists after JavaScript runs, a meaningful share of crawlers see an empty page.

How to check in thirty seconds: open your page, right-click, choose "View Page Source," and search that raw source for a sentence from your main content. If the sentence is not there, your content is being built client-side, and some crawlers will never see it.

Fixing the Rendering Gap

The fix is to deliver your important content in the initial HTML response rather than assembling it afterwards in the browser.

1. **Server-side rendering.** The server builds the full HTML and sends it complete. Best general answer for content that changes often.
2. **Static generation.** Pages are built ahead of time and served as finished files. Fastest and most reliable for content that changes rarely.
3. **Hybrid.** Critical content rendered server-side, non-essential interactive extras layered on afterwards. Usually the practical compromise.

What matters is not which framework you use but what arrives in that first response. Headings, body copy, structured data, and links should all be present before a single line of JavaScript executes.

Interactive extras — calculators, filters, live chat — can safely load afterwards. Your core answer cannot.

CHAPTER 04

Structure and Internal Linking

Site structure tells a machine what your business does, which pages matter most, and how topics relate. Done well it is invisible. Done badly it caps everything.

Two principles carry most of the weight:

1. **Shallow beats deep.** Any page worth ranking should be reachable within three clicks of the home page. Pages buried five levels down are treated as less important, because the structure says they are.
2. **Group by topic, not by date.** Organizing around subjects your buyers care about builds recognizable topical authority. Organizing by publication month builds nothing.

A useful test: could a stranger, looking only at your navigation, correctly describe what your business does and where it operates? If not, neither can a machine.

Hub and Spoke

The most reliable structure for a service business is hub and spoke. A hub page covers a topic broadly. Spoke pages cover individual questions in depth. Every spoke links back to its hub, and the hub links out to every spoke.

The effect is that authority pools at the hub and distributes to the spokes, and the relationship between the pages is unambiguous to anything reading the site.

PAGE TYPE	PURPOSE	LINKS TO
Hub	Broad coverage of one service	Every related spoke
Spoke	One specific question, answered fully	Its hub, plus sibling spokes
Location	One service in one place	Its hub, plus nearby locations

Keep anchor text descriptive. A link reading "our technical SEO services" carries meaning; one reading "click here" carries none.

URLs That Age Well

URLs are permanent in a way most site elements are not. Every change risks breaking links you do not control, so it is worth getting them right once.

- [ ] Short, readable, and made of real words rather than IDs or codes
- [ ] Lower case, with hyphens between words — never underscores or spaces
- [ ] No dates, unless the content is genuinely time-bound news
- [ ] No stop words padding the path out
- [ ] Structure that mirrors the site hierarchy
- [ ] Never changed without a permanent redirect from the old address

If you must change a URL, redirect the old one permanently and update your internal links to point at the new address directly. Chains of redirects leak both speed and authority.

CHAPTER 05

Speed That Actually Counts

Speed matters, but not every speed metric matters equally, and chasing a perfect score is a poor use of a budget.

Three measurements describe what a real visitor experiences:

METRIC	WHAT IT MEASURES	AIM FOR
LCP	When the main content appears	Under 2.5 seconds
INP	How fast the page reacts to a tap	Under 200 milliseconds
CLS	How much the layout jumps while loading	Under 0.1

Of the three, the first is usually the one costing you money — it is what a visitor experiences as "this site is slow," and it is most often caused by one oversized hero image.

The Fixes Worth Doing First

1. **Compress your images.** Almost always the single biggest win. Serve modern formats, size them for the space they occupy, and stop shipping 4000-pixel photos into 800-pixel slots.
2. **Reserve space for anything that loads late.** Images and embeds without declared dimensions shove the page around as they arrive, which is what a visitor experiences as a page fighting back.
3. **Audit your third-party scripts.** Chat widgets, heat maps, ad pixels and analytics accumulate quietly. Each one costs speed. Remove what nobody reads.
4. **Test on a real phone over mobile data,** not on office wifi on a desktop. That is the experience most of your visitors are actually having.

Do those four and you will capture most of the available benefit. Chasing the last few points of a lab score rarely changes what a visitor feels.

CHAPTER 06

Structured Data

Structured data is a block of machine-readable facts you add to a page, stating plainly what the page is about: this is a business, here is its address, here are its hours, this is an article, this person wrote it.

It does not change what a visitor sees. It changes how confidently a machine can describe you — which is exactly what an AI tool needs before it will name you in an answer.

The types that earn their keep for most businesses:

TYPE	USE IT ON	WHY IT MATTERS
Organization	Home page	Establishes who you are
LocalBusiness	Location pages	Address, hours, service area
Service	Service pages	What you actually do
FAQPage	Pages with real Q&A	Direct answers AI can lift
Article	Guides and posts	Author and date credibility

Getting Schema Right

Three rules prevent almost every structured data problem we encounter.

1. **It must match the visible page.** Marking up a review that does not appear on the page, or hours that contradict your contact page, is a violation and gets the markup ignored at best.
2. **Be consistent everywhere.** Your name, address, and phone number in schema must match your website, your Google Business Profile, and your directory listings exactly. "Street" in one place and "St." in another is a mismatch to a machine.
3. **Validate it after every deploy.** Schema breaks silently. A redesign or plugin update can strip it out without anybody noticing for months.

Structured data is cheap to add and cheap to maintain. It is also one of the few signals where being thorough puts you meaningfully ahead of competitors who have skipped it entirely.

CHAPTER 07

The Technical Audit

Run this quarterly. Everything here is checkable from a browser.

- [ ] Search *site:yourdomain.com* — does the count roughly match your real page count?
- [ ] Open robots.txt. Is anything important accidentally disallowed?
- [ ] View page source on your top page. Is the main content in the raw HTML?
- [ ] Is every important page reachable within three clicks of the home page?
- [ ] Does every page have exactly one H1, describing what the page is about?
- [ ] Do your URLs use real words, lower case, with hyphens?
- [ ] Load your top page on a phone over mobile data. Content in under three seconds?
- [ ] Does structured data exist, validate, and match what is on the page?
- [ ] Do NAP details match exactly across site, profile, and directories?
- [ ] Does every old URL you have changed still redirect permanently?

Priority Order When Everything Is Broken

Finding a dozen problems at once is normal. Fix them in this order, because each one gates the next.

PRIORITY	PROBLEM	WHY FIRST
1	Blocked or unindexed pages	Nothing else matters if bots cannot reach them
2	Content only rendered by JavaScript	Invisible content cannot rank or be cited
3	Broken redirects and error pages	Actively leaking authority right now
4	Missing or invalid structured data	Cheap to fix, disproportionate benefit
5	Speed and layout stability	Real, but rarely the binding constraint
6	Structure and internal linking	Compounding gain, slower to show

Teams routinely start at five because it produces a satisfying score. Start at one.

Keeping It Fixed

Technical health decays. Not because anybody is careless, but because sites change constantly and nothing warns you when a change breaks something invisible.

The usual culprits: a redesign that drops structured data, a plugin update that adds render-blocking scripts, a new section launched without internal links, a staging environment left crawlable, a URL changed without a redirect.

The defence is a routine, not a project:

1. **After every significant deploy,** re-check indexation and validate schema on a sample of key pages.
2. **Monthly,** review crawl errors and fix anything new.
3. **Quarterly,** run the full audit on the previous page.

Twenty minutes a month prevents the kind of problem that takes a quarter to unwind.

CHAPTER 08

Key Takeaways

1. Architecture sets the ceiling on what your content can achieve. Great writing on an unreadable site underperforms average writing on a well-built one.
2. Check crawlability before anything else. Content published behind a blocked path produces nothing.
3. If your main content only exists after JavaScript runs, a meaningful share of crawlers — including AI crawlers — see an empty page.
4. Keep important pages within three clicks, group by topic rather than date, and link with descriptive anchor text.
5. Compress your images before chasing any other speed fix. It is almost always the biggest single win.
6. Structured data must match the visible page, stay consistent across every listing, and be re-validated after each deploy.
7. Technical health decays silently. A quarterly audit costs less than one recovery.

ABOUT GOBIYA

Build on Solid Ground

Gobiya builds the technical foundation underneath durable organic growth — crawlability, rendering, structure, speed, and structured data. It is the least visible part of the work and the part that decides how far everything else can go. We have been doing this from Los Angeles since 2010.

Ready to see where you stand?

Schedule a free consultation and we'll walk through your baseline together.

323-744-1338 · gobiya.com