



Vertical Search Engine Architecture: Technical Audit Checklist

Phase 1: Content Boundary & Scope Definition

- **Domain Constraint:** Is the exact scope of the search engine defined and documented? (e.g., "Only technical documentation and API references").
- **Exclusion Rules:** Are irrelevant content types (marketing landing pages, legacy blog posts) actively excluded from the index?
- **Trust Signals:** Is there a clear visual indicator for the user explaining what data pool they are searching within?

Phase 2: Metadata Governance & Normalization

- **Required Fields:** Are core fields (Author, Publication Date, Document Type, Status) mandatory for all new entries?
- **Taxonomy Alignment:** Do tags and categories use a controlled vocabulary (preventing duplicates like "SEO", "Search Engine Optimization", "seo")?
- **Freshness Protocol:** Is there an automated system or routine to flag outdated content for review or archiving?

Phase 3: Entity Extraction & Schema Readiness

- **Schema Integration:** Is robust Schema.org markup ([Dataset](#), [TechArticle](#), [ItemPage](#)) implemented across the primary content types?
- **Entity Gap Audit:** Have you identified and mapped the industry-specific synonyms and acronyms your users actually search with?
- **RAG Compatibility:** Can your internal LLM routing cleanly extract plain-text answers from your structured data without hallucinating?

Phase 4: User Experience & Faceted Navigation



- [] **Context-Aware Filtering:** Can users filter results by metadata that influences their decision (e.g., "In Stock", "Approved Policy", "Jurisdiction")?
- [] **Snippet Clarity:** Do search results display a functional preview (abstract or technical specs) rather than just the first 150 characters of the page?
- [] **Zero-Result Logic:** Are zero-result queries tracked daily, and does the UI offer alternative pathways instead of a dead end?

Result: If you checked fewer than 8 boxes, your internal search is likely operating as a noisy directory rather than a precision vertical search engine.