

Streaming Search Readiness Checklist — One Page

Search is the highest-intent discovery surface — the one a viewer chooses to use. It does two jobs (find a known title; explore for a kind of thing) and can only find what your metadata describes. Build the keyword, typo, semantic, and hybrid layers, design for the TV remote, and never let a zero-result search be a dead end. Engine defaults and Google's structured-data rules move in 2026 — confirm live.

1 · THE TWO JOBS SEARCH MUST DO

- Known-item search** — the viewer typed a title they know. Optimize for speed, typo tolerance, and exact match.
- Exploratory search** — the viewer knows only a mood or "something like X." Needs meaning-matching; shades into recs.
- Serve both from one box** — a system tuned only for known-item returns blank screens to every mood query.
- Search reads your metadata** — it can only find what the catalog is described well enough to surface.

2 · THE SEARCH LAYERS (build up, don't skip)

- Keyword / lexical (BM25)** — inverted index + relevance ranking; the built-in default in Lucene/OpenSearch. Owns exact titles & names.
- Typo tolerance** — match by edit distance (Damerau-Levenshtein) so "strm" finds "Storm"; rank exact matches first.
- Semantic / vector** — embeddings + approximate nearest neighbor (HNSW) match meaning, not words. Owns moods & "like X."
- Hybrid + fusion** — run keyword and vector in parallel, fuse by rank (RRF), then re-rank the top with personalization.

THE ORDER OF OPERATIONS — KEYWORD + TYPO FIRST, THEN SEMANTIC, THEN HYBRID; NEVER A DEAD END

Search is the one discovery surface a viewer chooses to use, so failing it is the most expensive miss in the product — the viewer wanted to watch something and you could not connect them to it. Do it in order. First, ship competent keyword search: an inverted index with BM25 ranking (the built-in default in the engines most teams already run), made usable with typo tolerance so a misspelling still matches and autocomplete so the viewer barely has to type. Second, add semantic search when mood, paraphrase, and "something like X" queries matter — embeddings plus an approximate-nearest-neighbor index (HNSW) match meaning rather than words; delegate the embedding-model internals to your AI team and keep the product wiring here. Third, run both as hybrid search and fuse the two ranked lists by rank (Reciprocal Rank Fusion), then re-rank the top results with personalization — keyword owns the exact title, semantic owns the meaning, and hybrid serves both jobs from one box. Design for the television throughout: on a remote-driven 10-foot UI, typing is the enemy, so lean on autocomplete and voice, and remember voice produces the long exploratory queries only semantic search can answer. Finally, never let a zero-result search be a dead end — hand off to recommendations, feed successful searches back as intent signals, publish Schema.org structured data so external engines can find your titles, and track the zero-result and abandonment rates as retention levers. Engine defaults and Google's structured-data requirements move — re-verify before you plan against them.

3 · DESIGN FOR THE TELEVISION

- Autocomplete** — suggest titles after 1-2 characters; on a 10-foot UI it is the primary input, not a nicety.
- Voice search** — let viewers speak the query; the preferred alternative to slow D-pad on-screen text entry.
- Match voice to a semantic backend** — voice produces long, natural queries a keyword-only backend cannot answer.
- Forgive everything** — typos, partial words, plurals, and a leading "the" in titles.

4 · THE HANDOFF, OFF-PLATFORM & METRICS

- Zero-result handoff** — never a blank "No results"; fall back to closest matches, related genres, personalized rows.
- Feed the recommender** — a successful search is a strong intent signal; a failed one is the recommender's cue to step in.
- Off-platform discovery** — publish Schema.org VideoObject so Google and AI assistants can surface your titles.
- Track the right metrics** — zero-result rate and search-abandonment rate, plus NDCG / MRR for ranking quality.