

SpecForge

AI-Powered BIS Standards Recommendation Engine

The Problem



Manual Search Nightmare

Indian MSEs face an impossible task: 929 pages of BIS SP-21 documentation with 500+ standards

Current process: weeks of manual searching, domain expertise required, error-prone identification

No intelligent discovery system exists

Introducing SpecForge



Natural Language Input

Product description in plain
English



Millisecond Retrieval

Hybrid FAISS + BM25 search

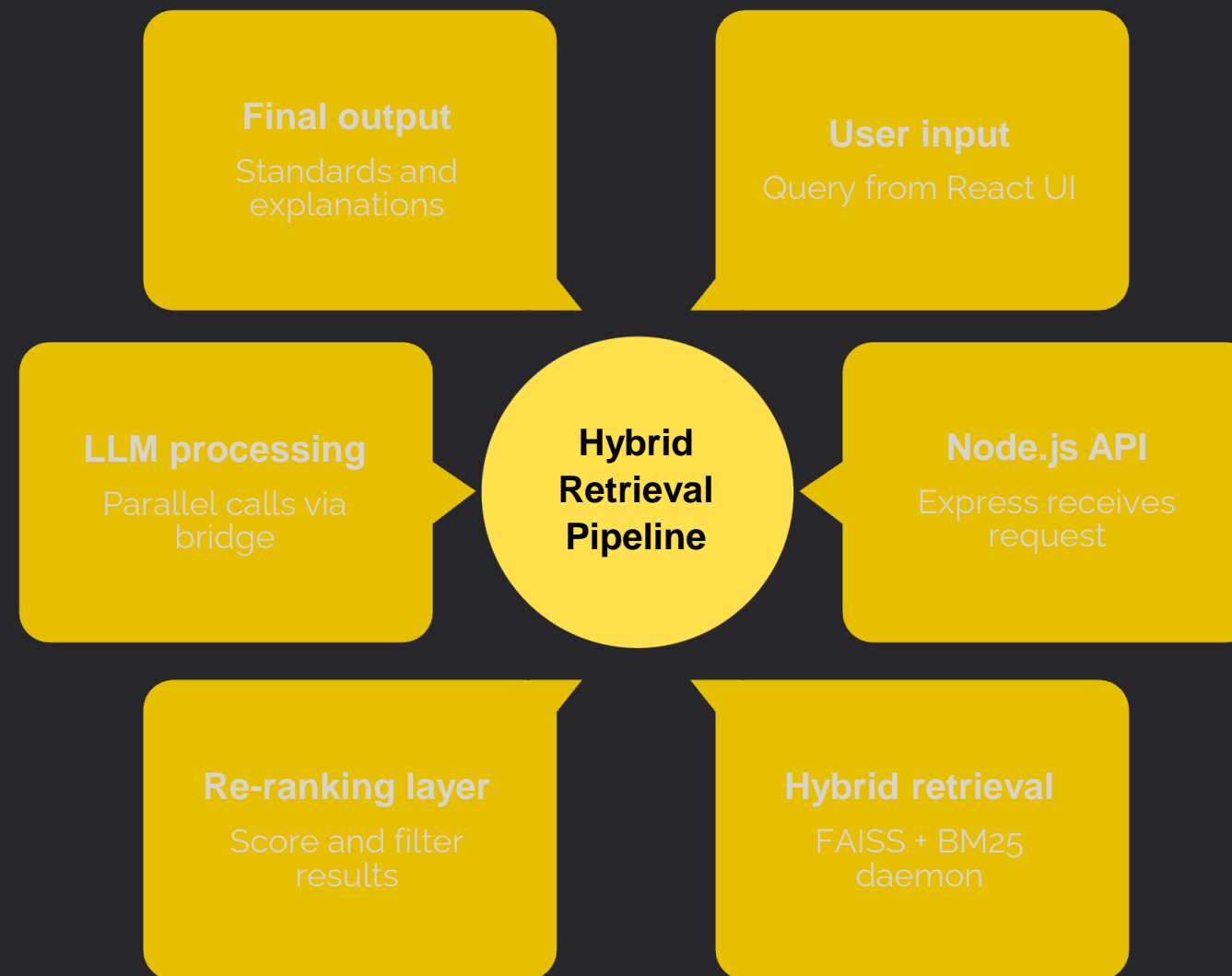


Top Standards + Explanations

Matched sections with AI-
generated context

From weeks of manual effort to milliseconds

System Architecture

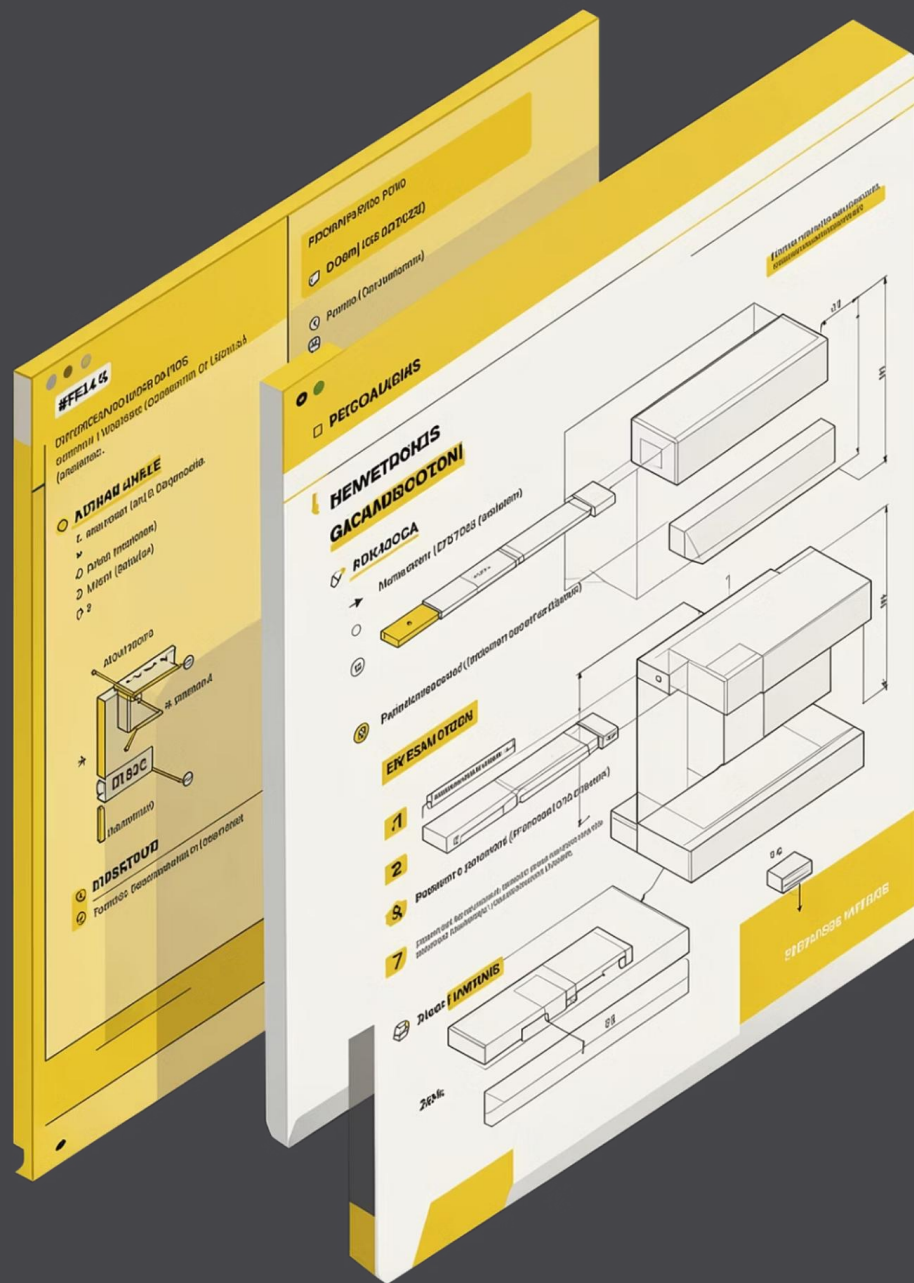


Backend Stack

- Python retrieval engine (FAISS + BM25)
- Persistent daemon eliminates cold start
- Node.js Express API

Frontend Stack

- React multi-page UI
- Python-Node bridge via subprocess
- Parallel LLM calls with Promise.allSettled



Chunking Strategy

01

PDF Parsing

929 pages → structured extraction

02

Standard Extraction

573 structured BIS standards identified

03

Chunk Creation

1,236 RAG-optimized chunks with overlap

04

Multi-Pass Validation

Fixed document structure issues automatically

Hybrid Retrieval Architecture



FAISS

Semantic search using all-MiniLM-L6-v2 embeddings



BM25

Keyword-based retrieval with weighted tokens



Fusion Score

$0.6 \times \text{Dense} + 0.4 \times \text{Sparse}$



Final Ranking

Top 3–5 standards with explanations

Advanced re-ranking: Keyword overlap, title match boosting, IS standard ID bonus, grade-aware logic (33/43/53 cement classification)



Demo: 33 Grade Cement

User Query

"33 Grade Ordinary
Portland Cement"

Retrieved Standards

1. IS 269:1989
2. IS 8112:1989
3. IS 12269:1987

Correct ranking • Instant response • AI
explanations

Performance Results

100%

Hit Rate @3

10/10 correct in top 3

1.0

MRR @5

Perfect ranking score

19ms

Average Latency

250× faster than requirement

Perfect accuracy with sub-20ms response time



Impact on MSEs

Time Savings

Weeks → seconds for compliance identification

Effort Reduction

Eliminates manual document searching

Accuracy Boost

Consistent, reliable standard matching

Scalability

Extensible to other BIS domains



Made by Engineers, for Engineers

Team

HackTheFuture

Built with Python, Node.js,
React

Powered by FAISS, Groq API,
PyMuPDF

Data Source

BIS SP-21 dataset

573 standards from 929 pages

Open Source

FAISS, HuggingFace, Groq