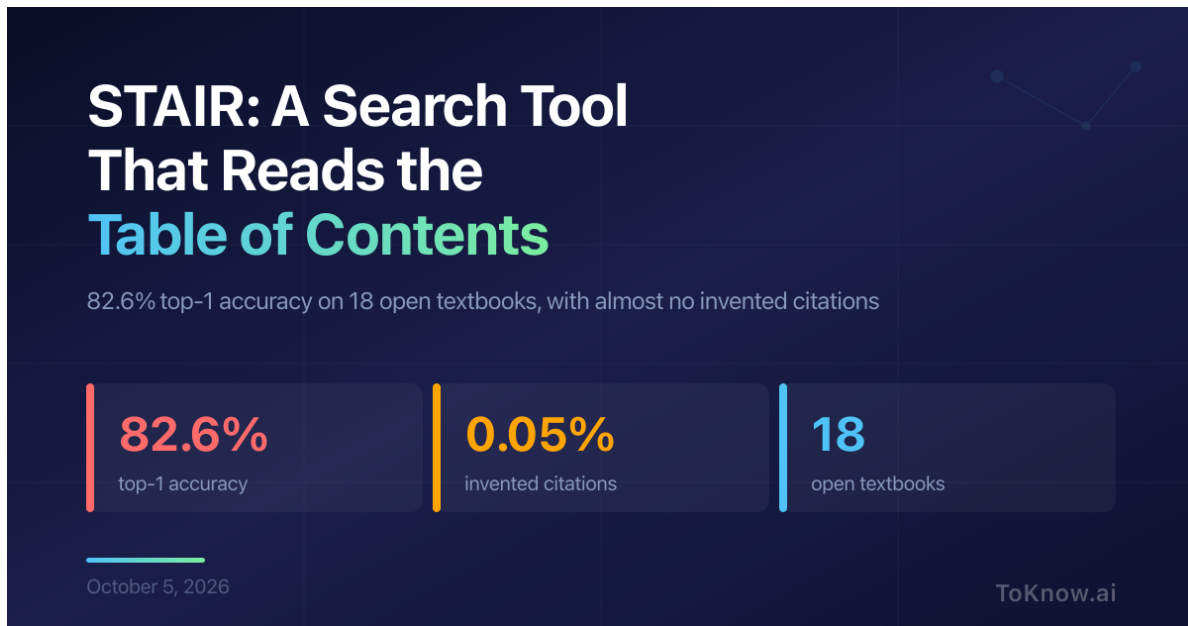


STAIR: A Search Tool That Reads the Table of Contents

Kabui, Charles

2026-10-05

 Read at ToKnow.ai



Researchers at IBM have built a search tool that answers questions about a long document by reading its table of contents first, the way a person flips to the right chapter. Most systems that ground an AI's answer in a document, called [retrieval-augmented generation](#), split the text into equal chunks and match them to a question, throwing away the document's structure. STAIR, short for STructure Aware Information Retriever, fine-tunes a small language model

to memorize the document, then asks it to pick the answering section from the full table of contents. Output is limited to section titles that exist, cutting invented citations to 0.05%. On [SearchTome](#), 18 open textbooks across six subjects, it named the right section first 82.6% of the time, against 76.9% for its closest baseline.

That matters for anyone searching a long, organized document: a repair manual, a contract, or a company knowledge base. Instead of an opaque chunk id like chunk_0417, it returns a real heading that a person can verify and an agent can trace back to the text. There is no separate vector database, because the index lives inside the model's weights. An independent developer who rebuilt it fit the whole thing into about five gigabytes of video memory, enough for one consumer graphics card.

STAIR only works when the document already has a table of contents, and the IBM team has not yet tested messy, unstructured sources like web pages or chat logs.

Read More: [SynthDocBench: The Middle of a Long Document Is Still a Blind Spot](#)

Sources:

- [STAIR: A novel dataset and LLM based retriever for document structure augmentation \(arXiv\)](#)
- [An open reimplementation of STAIR \(GitHub\)](#)
- [Hugging Face paper page](#)
- [CosmoNet: STAIR di IBM, il retriever RAG senza vector database](#)

***Disclaimer:** For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)*