Release Notes 9.8

?? Sep 2023, Apache Lucene™ 9.8 available

The Lucene PMC is pleased to announce the release of Apache Lucene 9.8

Apache Lucene is a high-performance, full-featured search engine library written entirely in Java. It is a technology suitable for nearly any application that requires structured search, full-text search, faceting, nearest-neighbor search on high-dimensionality vectors, spell correction or query suggestions.

This release contains numerous features, optimizations, and improvements, some of which are highlighted below. The release is available for immediate download at:

https://lucene.apache.org/core/downloads.html

Lucene 9.8 Release Highlights

Optimizations

- Faster computation of top-k hits on boolean queries. Lucene’s nightly benchmarks report a 20%-30% speedup for disjunctive queries and a 11% -13% speedup for conjunctive queries since Lucene 9.7. Disjunctive queries with many and/or high-frequency terms should see even higher speedups.
- Faster computation of top-k hits when sorting by field. Lucene’s nightly benchmarks report speedups between 7% and 33% since 9.7 depending on the type and cardinality of the field that is used for sorting.
- Faster counts on disjunctive queries by taking advantage of the new `LeafCollector.collect(DocIdStream)` API. Nightly benchmarks report a ~2.5x speedup since 9.7.
- Faster indexing of numeric doc values when index sorting is turned on.
- Expressions now evaluate all arguments in a fully lazy manner, which may provide significant speedups and throughput improvements for heavy expression users.

API Changes

- Move max vector dims limit to Codec (Mayya Sharipova)

New features

- Introduced `LeafCollector#finish`, a hook that runs after collection has finished running on a leaf.
- Add `KnnCollector` to `LeafReader` and `KnnVectorReader` so that custom collection of vector search results can be provided. The first custom collector provides `ToParentBlockJoin[Float|Byte]KnnVectorQuery` joining child vector documents with their parent documents.
- Add support for recursive graph bisection, also called bipartite graph partitioning, and often abbreviated BP, an algorithm for reordering doc IDs that results in more compact postings and faster queries, especially conjunctions.

Bug Fixes

- Fix HNSW graph search bug that potentially leaked unapproved docs
- Fix bug in `TermsEnum#seekCeil` on doc values terms enums that causes `IndexOutOfBoundsException`.

... plus a great number of other contributions!

Further details of changes are available in the change log available at: https://lucene.apache.org/core/9_8_0/changes/Changes.html.

Please report any feedback to the mailing lists (http://lucene.apache.org/core/discussion.html)

Note: The Apache Software Foundation now uses a content distribution network (CDN) for distributing releases.