

ReleaseNote40beta

MMM???? 2012, Apache Lucene, 4.0-beta available

The Lucene PMC is pleased to announce the release of Apache Lucene 4.0-beta

Apache Lucene is a high-performance, full-featured text search engine library written entirely in Java. It is a technology suitable for nearly any application that requires full-text search, especially cross-platform.

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

<http://lucene.apache.org/core/mirrors-core-latest-redir.html>

See the CHANGES.txt file included with the release for a full list of details.

Highlights of changes since 4.0-alpha:

- * IndexWriter.tryDeleteDocument can sometimes delete by document ID, for higher performance in some applications.
- * New experimental postings formats: BloomFilteringPostingsFormat uses a bloom filter to sometimes avoid disk seeks when looking up terms, DirectPostingsFormat holds all postings as simple byte[] and int[] for very fast performance at the cost of very high RAM consumption.
- * CJK analysis improvements: JapaneseIterationMarkCharFilter normalizes Japanese iteration marks, added unigram+bigram support to CJKBigramFilter.
- * Improvements to Scorer navigation API (Scorer.getChildren) to support all queries, useful for determining which portions of the query matched.
- * Analysis improvements: factories for creating Tokenizer, TokenFilter and CharFilter have been moved from Solr to Lucene's analysis module, less memory overhead for StandardTokenizer and Snowball filters.
- * Improved highlighting for multi-valued fields.
- * Various other API changes, optimizations and bug fixes.

Please read CHANGES.txt and MIGRATE.txt for a full list of new features and notes on upgrading. Particularly, the new apis are not compatible with previous version of Lucene, however, file format backwards compatibility is provided for indexes from the 3.0 series and the 4.0-alpha release.

This is a beta release for early adopters. The guarantee for this beta release is that the index format will be the 4.0 index format, supported through the 5.x series of Apache Lucene, unless there is a critical bug (e.g. that would cause index corruption) that would prevent this.

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