

ReleaseNote44

July 2013, Apache Lucene™ 4.4 available

The Lucene PMC is pleased to announce the release of Apache Lucene 4.4

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.

Lucene 4.4 Release Highlights:

- * New Replicator module: replicate index revisions between server and client. See <http://shaierera.blogspot.com/2013/05/the-replicator.html>
- * New AnalyzingInfixSuggester: finds suggestions based on matches to any tokens in the suggestion, not just based on pure prefix matching. See <http://blog.mikemccandless.com/2013/06/a-new-lucene-suggester-based-on-infix.html>
- * New PatternCaptureGroupTokenFilter: emit multiple tokens, one for each capture group in one or more Java regexes.
- * New Lucene Facet module features:
 - * Added dynamic (no taxonomy index used) numeric range faceting (see <http://blog.mikemccandless.com/2013/05/dynamic-faceting-with-lucene.html>)
 - * Arbitrary Querys are now allowed for per-dimension drill-down on DrillDownQuery and DrillSideways, to support future dynamic faceting.
 - * New FacetResult.mergeHierarchies: merge multiple FacetResult of the same dimension into a single one with the reconstructed hierarchy.
- * FST's Builder can now handle more than 2.1 billion "tail nodes" while building a minimal FST.
- * FieldCache Ints and Longs now use bit-packing to save memory. String fields have more efficient compression if there are many unique terms.
- * Improved compression for NumericDocValues for dates and fields with very small numbers of unique values.
- * New IndexWriter.hasUncommittedChanges(): returns true if there are changes that have not been committed.
- * multiValuedSeparator in PostingsHighlighter is now configurable, for cases where you want a different logical separator between field values.
- * NorwegianLightStemFilter and NorwegianMinimalStemFilter have been extended to handle "nynorsk".
- * New ScandinavianFoldingFilter and ScandinavianNormalizationFilter.
- * Easier compressed norms: Lucene42NormsFormat now takes an overhead parameter, allowing for values other than PackedInts.FASTEST.
- * Analyzer now has an additional tokenStream(String fieldName, String text) method, so wrapping by StringReader for common use is no longer needed.
- * New SimpleMergedSegmentWarmer: just ensures that data structures (terms, norms, docvalues, etc.) are initialized.
- * IndexWriter flushes segments to the compound file format by default.
- * Various bugfixes and optimizations since the 4.3.1 release.

Please read `CHANGES.txt` for a full list of new features.

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

Note: The Apache Software Foundation uses an extensive mirroring network for distributing releases. It is possible that the mirror you are using may not have replicated the release yet. If that is the case, please try another mirror. This also goes for Maven access.