

Apache Lucene - Overview

Table of contents

1 Apache Lucene.....	2
2 Lucene News.....	2
2.1 25 November 2009 - Lucene Java 3.0.0 available.....	2
2.2 6 November 2009 - Lucene Java 2.9.1 available.....	3
2.3 07 Oct. 2009 - Lucene at US ApacheCon November 2-6, 2009.....	3
2.4 25 September 2009 - Lucene Java 2.9.0 available.....	4
2.5 9 March 2009 - Lucene Java 2.4.1 available.....	5
2.6 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam.....	5
2.7 8 October 2008 - Lucene Java 2.4.0 available.....	6
2.8 06 May 2008 - Lucene Java 2.3.2 available.....	6
2.9 23 February 2008 - Lucene Java 2.3.1 available.....	7
2.10 24 January 2008 - Lucene Java 2.3.0 available.....	7
2.11 23 January 2008 - Lucene at ApacheCon Europe.....	8
2.12 24 December 2007 - Nightly Snapshots available in the Apache Maven Snapshot Repository.....	8
2.13 26 August 2007 - Lucene at ApacheCon Atlanta.....	8
2.14 19 June 2007 - Release 2.2 available.....	9
2.15 18 February 2007 - Lucene at ApacheCon Europe.....	9
2.16 17 February 2007 - Release 2.1 available.....	9
2.17 3 January 2007 - Nightly Source builds available.....	10

1. Apache Lucene

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.

Apache Lucene is an open source project available for free download. Please use the links on the left to access Lucene.

2. Lucene News

2.1. 25 November 2009 - Lucene Java 3.0.0 available

The new version is mostly a cleanup release without any new features. All deprecations targeted to be removed in version 3.0 were removed. If you are upgrading from version 2.9.1 of Lucene, you have to fix all deprecation warnings in your code base to be able to recompile against this version.

This is the first Lucene release with Java 5 as a minimum requirement. The API was cleaned up to make use of Java 5's generics, varargs, enums, and autoboxing. New users of Lucene are advised to use this version for new developments, because it has a clean, type safe new API. Upgrading users can now remove unnecessary casts and add generics to their code, too. If you have not upgraded your installation to Java 5, please read the file `JRE_VERSION_MIGRATION.txt` (*please note that this is not related to Lucene 3.0, it will also happen with any previous release when you upgrade your Java environment*).

Lucene 3.0 has some changes regarding compressed fields: 2.9 already deprecated compressed fields; support for them was removed now. Lucene 3.0 is still able to read indexes with compressed fields, but as soon as merges occur or the index is optimized, all compressed fields are decompressed and converted to `Field.Store.YES`. Because of this, indexes with compressed fields can suddenly get larger.

While we generally try and maintain full backwards compatibility between major versions, Lucene 3.0 has some minor breaks, mostly related to deprecation removal, pointed out in the 'Changes in backwards compatibility policy' section of `CHANGES.txt`. Notable are:

- `IndexReader.open(Directory)` now opens in read-only mode per default (this method was deprecated because of that in 2.9). The same occurs to `IndexSearcher`.
- Already started in 2.9, core `TokenStreams` are now made final to enforce the decorator pattern.
- If you interrupt an `IndexWriter` merge thread, `IndexWriter` now throws an unchecked `ThreadInterruptedException` that extends `RuntimeException`

and clears the interrupt status.

See CHANGES for details.

Binary and source distributions are available here. Maven artifacts are available here.

2.2. 6 November 2009 - Lucene Java 2.9.1 available

This release fixes bugs from 2.9.0, including one serious bug whereby BooleanQuery could silently fail to retrieve certain matching documents.

There are also some minor API changes, including a Version parameter added to QueryParser and contrib Analyzers, so that version dependent defaults are consistent across classes, as well as un-deprecating of certain methods (we were too zealous in a few cases!).

Otherwise the changes are all bug fixes and documentation improvements.

This release is fully compatible with 2.9.0. We strongly recommend upgrading to 2.9.1 if you are using 2.9.0. Furthermore, because some additional APIs were deprecated in 2.9.1, to ensure a clean ("JAR drop in") upgrade to 3.0 you'll need to ensure your code compiles against 2.9.1 without deprecation warnings.

See CHANGES for details.

Binary and source distributions are available here.

Maven artifacts are available here.

2.3. 07 Oct. 2009 - Lucene at US ApacheCon November 2-6, 2009



ApacheCon US is once again in the Bay Area and Lucene is coming along for the ride! The Lucene community has planned two full days of talks, plus a meetup and the usual bevy of training. With a well-balanced mix of first time and veteran ApacheCon speakers, the Lucene track at ApacheCon US promises to have something for everyone. Be sure not to miss:

Training:

- Lucene Boot Camp - A two day training session, Nov. 2nd & 3rd
- Solr Day - A one day training session, Nov. 2nd

Thursday, Nov. 5th

- Introduction to the Lucene Ecosystem - Grant Ingersoll @ 9:00
- Lucene Basics and New Features - Michael Busch @ 10:00
- Apache Solr: Out of the Box - Chris Hostetter @ 14:00
- Introduction to Nutch - Andrzej Bialecki @ 15:00
- Lucene and Solr Performance Tuning - Mark Miller @ 16:30

Friday, Nov. 6th (these sessions are all available via FREE ApacheCon US live video streams!)

- Implementing an Information Retrieval Framework for an Organizational Repository - Sithu D Sudarsan @ 9:00
- Apache Mahout - Going from raw data to Information - Isabel Drost @ 10:00
- MIME Magic with Apache Tika - Jukka Zitting @ 11:30
- Building Intelligent Search Applications with the Lucene Ecosystem - Ted Dunning @ 14:00
- Realtime Search - Jason Rutherglen @ 15:00

2.4. 25 September 2009 - Lucene Java 2.9.0 available

This release has many improvements since release 2.4.1, including:

- Per segment searching and caching (can lead to much faster reopen among other things)
- Near real-time search capabilities added to IndexWriter
- New Query types
- Smarter, more scalable multi-term queries (wildcard, range, etc)
- A freshly optimized Collector/Scorer API
- Improved Unicode support and the addition of Collation contrib
- A new Attribute based TokenStream API
- A new QueryParser framework in contrib with a core QueryParser replacement impl included.
- Scoring is now optional when sorting by Field, or using a custom Collector, gaining sizable performance when scores are not required.
- New analyzers (PersianAnalyzer, ArabicAnalyzer, SmartChineseAnalyzer)
- New fast-vector-highlighter for large documents
- Lucene now includes high-performance handling of numeric fields. Such fields are indexed with a trie structure, enabling simple to use and much faster numeric range searching without having to externally pre-process numeric values into textual values.

See CHANGES for details.

While we generally try and maintain full backwards compatibility between major versions, Lucene 2.9 has a variety of breaks that are spelled out in the 'Changes in backwards

compatibility policy' section of CHANGES. We recommend that you recompile your application with Lucene 2.9 rather than attempting to “drop” it in. This will alert you to any issues you may have to fix if you are affected by one of the backward compatibility breaks.

Binary and source distributions are available [here](#).

Maven artifacts are available [here](#).

2.5. 9 March 2009 - Lucene Java 2.4.1 available

This release contains fixes for bugs found in 2.4.0, including one data loss bug (LUCENE-1452) where in certain situations binary fields would be truncated to 0 bytes.

See CHANGES for details.

2.4.1 does not contain any new features, API or file format changes, which makes it fully compatible with 2.4.0.

Binary and source distributions are available [here](#).

Maven artifacts are available [here](#).

2.6. 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam



Lucene will be extremely well represented at ApacheCon EU 2009 in Amsterdam, Netherlands this March 23-27, 2009:

- Lucene Boot Camp - A two day training session, March 23 & 24th
- Solr Boot Camp - A one day training session, March 24th
- Introducing Apache Mahout - Grant Ingersoll. March 25th @ 10:30
- Lucene/Solr Case Studies - Erik Hatcher. March 25th @ 11:30
- Advanced Indexing Techniques with Apache Lucene - Michael Busch. March 25th @ 14:00
- Apache Solr - A Case Study - Uri Boness. March 26th @ 17:30

- Best of breed - httpd, forrest, solr and droids - Thorsten Scherler. March 27th @ 17:30
- Apache Droids - an intelligent standalone robot framework - Thorsten Scherler. March 26th @ 15:00

2.7. 8 October 2008 - Lucene Java 2.4.0 available

This release has many improvements since release 2.3.2, including:

- New InstantiatedIndex (contrib/instantiated): RAM-based index that enables much faster searching than RAMDirectory.
- New IndexWriter constructors now default autoCommit to false.
- New commit() method in IndexWriter lets you control when changes are made visible and permanent in the index.
- A machine or OS crash, or power loss, while IndexWriter is writing to an index will no longer corrupt the index.
- TimeLimitedCollector adds timeout to searches.
- Delete documents by Query in IndexWriter.
- Pure boolean indexing (no frequency, positions nor payloads are indexed) using Field.setOmitTf().
- A new Directory implementation, NIOFSDirectory, using java.nio's APIs to allow multiple threads to read from the same open file without locking.
- IndexWriter.expungeDeletes() reclaims disk space from deleted documents by merging away segments that have deletions.
- All filters now return a DocIdSet instead of java.util.BitSet, making filters more efficient and flexible.
- Searching with a Filter is more efficient: now the filter is applied to a document before scoring is done.
- IndexReader can be opened with new readOnly=true mode, which gives better performance in a multi-threaded environment.

See CHANGES for details.

Lucene 2.4.0 includes index format changes that are not readable by older versions of Lucene. Lucene 2.4.0 can both read and update older Lucene indexes. Adding to an index with an older format will cause it to be converted to the newer format.

Binary and source distributions are available here.

Maven artifacts are available here.

2.8. 06 May 2008 - Lucene Java 2.3.2 available

This release contains fixes for bugs found in 2.3.1.

See CHANGES.txt for a detailed listing of changes.

2.3.2 does not contain any new features, API or file format changes, which makes it fully compatible to 2.3.0 and 2.3.1.

Binary and source distributions are available here.

Maven artifacts are available here.

2.9. 23 February 2008 - Lucene Java 2.3.1 available

This release contains fixes for serious bugs in 2.3.0 that could cause index corruptions in autoCommit=false mode or in cases where multiple threads are adding documents where some have term-vector enabled fields and some don't. The autoCommit option was added to IndexWriter with release 2.2.0. If not explicitly set to false, the IndexWriter runs in autoCommit=true mode.

See CHANGES.txt for a detailed listing of changes.

2.3.1 does not contain any new features, API or file format changes, which makes it fully compatible to 2.3.0.

We would like to encourage everyone who is currently using Lucene Java 2.3.0 to upgrade to 2.3.1 to prevent possible index corruptions!

Binary and source distributions are available here.

Maven artifacts are available here.

2.10. 24 January 2008 - Lucene Java 2.3.0 available

This release has many improvements since release 2.2, including:

- Significantly improved indexing performance
- Segment merging in background threads
- Refreshable IndexReaders
- Faster StandardAnalyzer and improved Token API
- TermVectorMapper to customize how term vectors are loaded
- Live backups (without pausing indexing) with SnapshotDeletionPolicy
- CheckIndex tool to test and recover a corrupt index
- Pluggable MergePolicy and MergeScheduler
- "Partial" optimize(int maxNumSegments) method
- New contrib module for working with Wikipedia content

In addition Lucene 2.3.0 has many performance improvements, bug fixes, etc. See CHANGES.txt for details.

Lucene 2.3.0 includes index format changes that are not readable by older versions of Lucene. Lucene 2.3.0 can both read and update older Lucene indexes. Adding to an index with an older format will cause it to be converted to the newer format.

Binary and source distributions are available here.

Maven artifacts are available here.

2.11. 23 January 2008 - Lucene at ApacheCon Europe

Lucene projects will be well represented at ApacheCon Europe in Amsterdam this year. Please join us at one or more of the following sessions:

- April 7: Lucene Boot Camp training by Grant Ingersoll.
- April 8: Solr Boot Camp training by Erik Hatcher.
- April 8: Mastering and extending Apache Nutch training by Sami Siren.
- April 10: Lucene/Solr Case Studies by Erik Hatcher
- April 11: Apache Lucene Performance by Grant Ingersoll

2.12. 24 December 2007 - Nightly Snapshots available in the Apache Maven Snapshot Repository

We are now publishing nightly artifacts to the Maven Snapshot Repository. The current version is '2.3-SNAPSHOT'.

The artifacts include:

- Binary jars
- Sources
- Javadocs

You can find separate artifacts for the core, demo, and the different contrib modules.

Merry Christmas!

2.13. 26 August 2007 - Lucene at ApacheCon Atlanta

Lucene will once again be well represented at ApacheCon USA in Atlanta this November 12-16, 2007.

The following talks and trainings are scheduled for this year's conference:

- November 12: Lucene Boot Camp by Grant Ingersoll. An all-day training focusing on

- getting started with Lucene.
- November 16, 9:00 am: Apache Solr out of the Box by Chris Hostetter. Introduction to Solr.
 - November 16, 10:00 am: Building a Vertical Search Site using Apache Software by Ken Krugler. Will cover many Lucene-based projects.
 - November 16, 3:00 pm: Apache Lucene Performance by Grant Ingersoll. Tips and techniques for improving Lucene performance.
 - November 16, 4:00 pm: Advanced Indexing Techniques with Apache Lucene by Michael Busch. Information on payloads and advanced indexing techniques.

2.14. 19 June 2007 - Release 2.2 available

This release has many improvements since release 2.1. New major features:

- "Point-in-time" searching over NFS
- Payloads
- Function queries
- New API for pre-analyzed fields

In addition Lucene 2.2 has many performance improvements, bug fixes, etc. See CHANGES.txt for details.

Lucene 2.2 includes index format changes that are not readable by older versions of Lucene. Lucene 2.2 can both read and update older Lucene indexes. Adding to an index with an older format will cause it to be converted to the newer format.

Binary and source distributions are available here.

2.15. 18 February 2007 - Lucene at ApacheCon Europe

Lucene Java and related Lucene projects will have extensive representation at ApacheCon Europe in Amsterdam this year. For the 2007 session, Yonik Seeley will be giving the Full-Text Search with Lucene talk at 10:30 am on May 2nd. Immediately following, Grant Ingersoll will be presenting Advanced Lucene at 11:30. Grant will also be leading a full day tutorial session on May 1st titled Lucene Boot Camp.

Lucene related talks include Solr committer Bertrand Delacrétaiz's talk titled Beyond full-text searches with Solr and Lucene and Hadoop committer Owen O'Malley's Introduction to Hadoop.

Registration is now open on the ApacheCon website.

2.16. 17 February 2007 - Release 2.1 available

This release has many improvements since release 2.0, including new features, performance improvements, bug fixes, etc. See CHANGES.txt for details.

Lucene 2.1 includes index format changes that are not readable by older versions of Lucene. Lucene 2.1 can both read and update older Lucene indexes. Adding to an index with an older format will cause it to be converted to the newer format.

Binary and source distributions are available here.

2.17.3 January 2007 - Nightly Source builds available

Nightly source builds of the current development version of Lucene are now available at <http://people.apache.org/builds/lucene/java/nightly/>. Files are named lucene-DATE-src.tar.gz where DATE is the date of the build.