

Apache Lucene - Overview

Table of contents

1 Apache Lucene.....	2
2 Lucene News.....	2
2.1 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam.....	2
2.2 8 October 2008 - Lucene Java 2.4.0 available.....	2
2.3 06 May 2008 - Lucene Java 2.3.2 available.....	3
2.4 23 February 2008 - Lucene Java 2.3.1 available.....	4
2.5 24 January 2008 - Lucene Java 2.3.0 available.....	4
2.6 23 January 2008 - Lucene at ApacheCon Europe.....	5
2.7 24 December 2007 - Nightly Snapshots available in the Apache Maven Snapshot Repository.....	5
2.8 26 August 2007 - Lucene at ApacheCon Atlanta.....	5
2.9 19 June 2007 - Release 2.2 available.....	6
2.10 18 February 2007 - Lucene at ApacheCon Europe.....	6
2.11 17 February 2007 - Release 2.1 available.....	6
2.12 3 January 2007 - Nightly Source builds available.....	7

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. 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

2.2. 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.

2.3.06 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.

2.4. 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](#).

2.5. 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.

2.6. 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

2.7. 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!

2.8. 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.

2.9. 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.

2.10. 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.

2.11. 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.

2.12. 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.