

Welcome to Lucene!

Table of contents

1 What Is Lucene?.....	2
2 News.....	2
2.1 10 November 2009 - Solr 1.4 Released.....	2
2.2 6 November 2009 - Lucene Java 2.9.1 available.....	3
2.3 25 September 2009 - Lucene Java 2.9.0 available.....	3
2.4 14 August 2009 - Lucene at US ApacheCon.....	4
2.5 25 June 2009 - Apache Open Relevance Kickoff.....	5
2.6 07 April 2009 - Apache Mahout 0.1 released.....	5
2.7 9 March 2009 - Lucene Java 2.4.1 available.....	5
2.8 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam.....	6
2.9 19 January 2009 - PyLucene joins the Lucene TLP.....	6
2.10 8 October 2008 - Lucene Java 2.4.0 available.....	6
2.11 15 September 2008 - Solr 1.3.0 Available.....	6

1. What Is Lucene?

The Apache Lucene project develops open-source search software, including:

- [Lucene Java](#), our flagship sub-project, provides Java-based indexing and search technology, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities.
- [Droids](#) is an intelligent robot crawling framework currently in incubation.
- [Lucene.Net](#) is a source code, class-per-class, API-per-API and algorithmic port of the [Lucene Java](#) search engine to the C# and .NET platform utilizing Microsoft .NET Framework. Lucene.Net is currently under incubation.
- [Lucy](#) is a loose C port of [Lucene Java](#), with Perl and Ruby bindings.
- [Mahout](#) is a subproject with the goal of creating a suite of scalable machine learning libraries.
- [Nutch](#) builds on [Lucene Java](#) to provide web search application software.
- [Open Relevance Project](#) is a new subproject with the aim of collecting and distributing free materials for relevance testing and performance.
- [PyLucene](#) is a Python port of the the [Lucene Java](#) project.
- [Solr](#) is a high performance search server built using [Lucene Java](#), with XML/HTTP and JSON/Python/Ruby APIs, hit highlighting, faceted search, caching, replication, and a web admin interface.
- [Tika](#) is a toolkit for detecting and extracting metadata and structured text content from various documents using existing parser libraries.

2. News

2.1. 10 November 2009 - Solr 1.4 Released

Solr 1.4 has been released and is now available for public [download](#)! New Solr 1.4 features include

- Major performance enhancements in indexing, searching, and faceting
- Revamped all-Java index replication that's simple to configure and can replicate config files
- Greatly improved database integration via the DataImportHandler
- Rich document processing (Word, PDF, HTML) via Apache Tika
- Dynamic search results clustering via Carrot2
- Multi-select faceting (support for multiple items in a single category to be selected)
- Many powerful query enhancements, including ranges over arbitrary functions, nested queries of different syntaxes
- Many other plugins including Terms for auto-suggest, Statistics, TermVectors,

Deduplication

See the [release notes](#) for more details.

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. 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](#).

2.4. 14 August 2009 - Lucene at US ApacheCon

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

- [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

2.5. 25 June 2009 - Apache Open Relevance Kickoff

The Apache Lucene PMC has officially voted to add the Open Relevance Project (ORP) as a Lucene subproject. ORP's main goal is to build out collections, judgments and queries in an open environment to make it easier for Lucene developers and users to do relevance testing, much like one would get if using [TREC](#) or other evaluation conferences.

See <http://lucene.apache.org/openrelevance> for more info

2.6. 07 April 2009 - Apache Mahout 0.1 released

The Apache Lucene project is pleased to announce the release of Apache Mahout 0.1. Apache Mahout is a subproject of Apache Lucene with the goal of delivering scalable machine learning algorithm implementations under the Apache license. The first public release includes implementations for clustering, classification, collaborative filtering and evolutionary programming.

Highlights include:

- Taste Collaborative Filtering
- Several distributed clustering implementations: k-Means, Fuzzy k-Means, Dirchlet, Mean-Shift and Canopy
- Distributed Naive Bayes and Complementary Naive Bayes classification implementations
- Distributed fitness function implementation for the Watchmaker evolutionary programming library
- Most implementations are built on top of Apache Hadoop (<http://hadoop.apache.org>) for scalability

More info is available on the Mahout [website](#).

2.7. 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](#).

2.8. 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

2.9. 19 January 2009 - PyLucene joins the Lucene TLP

[PyLucene](#), the Python based port of Lucene is now an official Lucene subproject.

2.10. 8 October 2008 - Lucene Java 2.4.0 available

Lucene 2.4.0 is available for public download. This version contains many enhancements and bug fixes. See [CHANGES](#) for details.

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

Maven artifacts are available [here](#).

2.11. 15 September 2008 - Solr 1.3.0 Available

Solr 1.3.0 is available for public download. This version contains many enhancements and

Welcome to Lucene!

bug fixes, including distributed search capabilities, Lucene 2.3.x performance improvements and many others.

See the [release notes](#) for more details. Download is available from a [Apache Mirror](#).