

# Welcome to Solr

## Table of contents

1 What Is Solr?.....	3
2 Get Started.....	3
3 News.....	4
3.1 18 November 2011 - 2nd edition of the first book on Solr published!.....	4
3.2 26 October 2011 - Java 7u1 fixes index corruption and crash bugs in Apache Lucene Core and Apache Solr.....	5
3.3 14 September 2011 - Solr 3.4.0 Released.....	5
3.4 28 July 2011 - WARNING: Index corruption and crashes in Apache Lucene Core / Apache Solr with Java 7.....	6
3.5 July 2011 - Solr 3.3 Released.....	7
3.6 May 2011 - Solr 3.2 Released.....	7
3.7 March 2011 - Solr 3.1 Released.....	8
3.8 25 June 2010 - Solr 1.4.1 Released.....	8
3.9 7 May 2010 - Apache Lucene Eurocon 2010 Coming to Prague May 18-21.....	8
3.10 10 November 2009 - Solr 1.4 Released.....	9
3.11 20 August 2009 - Solr's first book is published!.....	10
3.12 18 August 2009 - Lucene at US ApacheCon.....	11
3.13 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam.....	11
3.14 19 December 2008 - Solr Logo Contest Results.....	12
3.15 03 October 2008 - Solr Logo Contest.....	12
3.16 15 September 2008 - Solr 1.3.0 Available.....	12
3.17 28 August 2008 - Lucene/Solr at ApacheCon New Orleans.....	12
3.18 03 September 2007 - Lucene at ApacheCon Atlanta.....	13

3.19 06 June 2007: Release 1.2 available.....	13
3.20 17 January 2007: Solr graduates from Incubator.....	13
3.21 22 December 2006: Release 1.1.0 available.....	13
3.22 15 August 2006: Solr at ApacheCon US.....	14
3.23 21 April 2006: Solr at ApacheCon.....	14
3.24 21 February 2006: nightly builds.....	14
3.25 17 January 2006: Solr Joins Apache Incubator.....	14

## 1 What Is Solr?

Solr is the popular, blazing fast open source enterprise search platform from the Apache Lucene project. Its major features include powerful full-text search, hit highlighting, faceted search, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search. Solr is highly scalable, providing distributed search and index replication, and it powers the search and navigation features of many of the world's largest internet sites.

Solr is written in Java and runs as a standalone full-text search server within a servlet container such as [Tomcat](#). Solr uses the [Lucene Java](#) search library at its core for full-text indexing and search, and has REST-like HTTP/XML and JSON APIs that make it easy to use from virtually any programming language. Solr's powerful external configuration allows it to be tailored to almost any type of application without Java coding, and it has an extensive plugin architecture when more advanced customization is required.

See the complete [feature list](#) for more details.

For more information about Solr, please see the [Solr wiki](#).

## 2 Get Started

- Download Solr [here](#)
- Check out the [tutorial](#)
- Read the [Solr wiki](#) to learn more
- Join the [community](#)
- [Give Back!](#) (Optional, but encouraged!)

### 3 News

#### 3.1 18 November 2011 - 2nd edition of the first book on Solr published!



David Smiley and Eric Pugh proudly announce the second edition of the first book on Solr, ["Apache Solr 3 Enterprise Search Server"](#) from Packt Publishing.

Apache Solr 3 Enterprise Search Server is a comprehensive reference guide for nearly every feature has to offer. Through using a large set of metadata about artists, releases, and tracks courtesy of the MusicBrainz.org project, you will have a testing ground for learning Solr. You'll learn how to design a schema, use appropriate text analysis and then how to import this data in various ways. Next, you'll learn how to search this data, how to use advanced relevancy tuning, and how to enhance standard search results with highlighting, faceting, query auto-complete, and other features. The book, supported with working code examples in various languages, shows how to use a wide selection of Solr integration client libraries, frameworks and other software like web crawlers. The book wraps up with deployment considerations, tuning Solr performance, and scaling Solr to multiple machines.

This edition naturally covers the latest features in Solr as of version 3.4 like Result Grouping and Geospatial, but this is not a small update to the first book. No chapter was untouched — some were revamped significantly and the content was expanded by about 25% by page count. Each chapter has a tip in the introduction that advises readers in a hurry on what parts should be read now or later. Finally, it includes a 2-page parameter quick-reference appendix that you will surely find useful printed on your desk.

You can find further information at [the publisher's site](#) and at [the authors' site](#), including a free chapter and search parameter quick-reference sheet (the appendix).

### 3.2 26 October 2011 - Java 7u1 fixes index corruption and crash bugs in Apache Lucene Core and Apache Solr

Oracle released [Java 7u1](#) on October 19. According to the release notes and tests done by the Lucene committers, all bugs reported on July 28 are fixed in this release, so code using Porter stemmer no longer crashes with SIGSEGV. We were not able to experience any index corruption anymore, so it is safe to use Java 7u1 with Lucene Core and Solr.

On the same day, Oracle released [Java 6u29](#) fixing the same problems occurring with Java 6, if the JVM switches `-XX:+AggressiveOpts` or `-XX:+OptimizeStringConcat` were used. Of course, you should **not** use experimental JVM options like `-XX:+AggressiveOpts` in production environments! We recommend everybody to upgrade to this latest version 6u29.

In case you upgrade to Java 7, remember that you may have to reindex, as the unicode version shipped with Java 7 changed and tokenization behaves differently (e.g. lowercasing). For more information, read `JRE_VERSION_MIGRATION.txt` in your distribution package!

### 3.3 14 September 2011 - Solr 3.4.0 Released

The Lucene PMC is pleased to announce the release of [Apache Solr 3.4.0](#)!

Solr's version number was synced with Lucene following the Lucene/Solr merge, so Solr 3.4.0 contains Lucene 3.4.0.

If you are already using Apache Solr 3.1, 3.2 or 3.3, we strongly recommend you upgrade to 3.4.0 because of the index corruption bug on OS or computer crash or power loss ([LUCENE-3418](#)), now fixed in 3.4.0.

Solr 3.4.0 release highlights include

- Bug fixes and improvements from Apache Lucene 3.4.0, including a major bug ([LUCENE-3418](#)) whereby a Lucene index could easily become corrupted if the OS or computer crashed or lost power.
- SolrJ client can now parse grouped and range facets results ([SOLR-2523](#)).
- A new `XsltUpdateRequestHandler` allows posting XML that's transformed by a provided XSLT into a valid Solr document ([SOLR-2630](#)).
- Post-group faceting option (`group.truncate`) can now compute facet counts for only the highest ranking documents per-group. ([SOLR-2665](#)).

- Add `commitWithin` update request parameter to all update handlers that were previously missing it. This tells Solr to commit the change within the specified amount of time ([SOLR-2540](#)).
- You can now specify `NIOFSDirectory` ([SOLR-2670](#)).
- New parameter `hl.phraseLimit` speeds up `FastVectorHighlighter` ([LUCENE-3234](#)).
- The query cache and filter cache can now be disabled per request. See [this wiki page](#) ([SOLR-2429](#)).
- Improved memory usage, build time, and performance of `SynonymFilterFactory` ([LUCENE-3233](#)).
- Added `omitPositions` to the schema, so you can omit position information while still indexing term frequencies ([LUCENE-2048](#)).
- Various fixes for multi-threaded `DataImportHandler`.

See the [release notes](#) for a more complete list of all the new features, improvements, and bugfixes.

### 3.4 28 July 2011 - WARNING: Index corruption and crashes in Apache Lucene Core / Apache Solr with Java 7

Oracle released [Java 7](#) today. Unfortunately it contains hotspot compiler optimizations, which miscompile some loops. This can affect code of several Apache projects. Sometimes JVMs only crash, but in several cases, results calculated can be incorrect, leading to bugs in applications (see Hotspot bugs [7070134](#), [7044738](#), [7068051](#)).

**Apache Lucene Core** and **Apache Solr** are two Apache projects, which are affected by these bugs, namely all versions released until today. Solr users with the default configuration will have Java crashing with SIGSEGV as soon as they start to index documents, as one affected part is the well-known Porter stemmer (see [LUCENE-3335](#)). Other loops in Lucene may be miscompiled, too, leading to index corruption (especially on Lucene trunk with pulsing codec; other loops may be affected, too - [LUCENE-3346](#)).

These problems were detected only 5 days before the official Java 7 release, so Oracle had no time to fix those bugs, affecting also many more applications. In response to our questions, they proposed to include the fixes into service release u2 (eventually into service release u1, see [this mail](#)). **This means you cannot use Apache Lucene/Solr with Java 7 releases before Update 2!** If you do, please don't open bug reports, it is not the committers' fault! At least disable loop optimizations using the `-XX:-UseLoopPredicate` JVM option to not risk index corruptions.

*Please note:* Also Java 6 users are affected, if they use one of those JVM options, which are **not** enabled by default: `-XX:+OptimizeStringConcat` or `-XX:+AggressiveOpts`.

It is strongly recommended not to use any hotspot optimization switches in any Java version without extensive testing!

In case you upgrade to Java 7, remember that you may have to reindex, as the unicode version shipped with Java 7 changed and tokenization behaves differently (e.g. lowercasing). For more information, read `JRE_VERSION_MIGRATION.txt` in your distribution package!

### 3.5 July 2011 - Solr 3.3 Released

The Lucene PMC is pleased to announce the release of [Apache Solr 3.3](#)!

Solr's version number was synced with Lucene following the Lucene/Solr merge, so Solr 3.3 contains Lucene 3.3.

Solr 3.3 release highlights include

- Grouping / Field Collapsing
- A new, automaton-based suggest/autocomplete implementation offering an order of magnitude smaller RAM consumption.
- KStemFilterFactory, an optimized implementation of a less aggressive stemmer for English.
- Solr defaults to a new, more efficient merge policy (TieredMergePolicy). See <http://s.apache.org/merging> for more information.
- Important bugfixes, including extremely high RAM usage in spellchecking.
- Bugfixes and improvements from Apache Lucene 3.3

See the [release notes](#) for a more complete list of all the new features, improvements, and bugfixes.

### 3.6 May 2011 - Solr 3.2 Released

The Lucene PMC is pleased to announce the release of [Apache Solr 3.2](#)!

Solr's version number was synced with Lucene following the Lucene/Solr merge, so Solr 3.2 contains Lucene 3.2. Solr 3.2 is the first release after Solr 3.1.

Solr 3.2 release highlights include

- Ability to specify `overwrite` and `commitWithin` as request parameters when using the JSON update format
- `TermQParserPlugin`, useful when generating filter queries from terms returned from field faceting or the terms component.
- `DebugComponent` now supports using a `NamedList` to model `Explanation` objects in its responses instead of `Explanation.toString`
- Improvements to the UIMA and Carrot2 integrations

- Bugfixes and improvements from Apache Lucene 3.2

See the [release notes](#) for a more complete list of all the new features, improvements, and bugfixes.

### 3.7 March 2011 - Solr 3.1 Released

Solr 3.1 has been released and is now available for public [download](#)! New Solr 3.1 features include

- Improved geospatial support
- Sorting by function queries
- Range faceting on all numeric fields
- Example Velocity driven search UI at <http://localhost:8983/solr/browse>
- A new termvector-based highlighter
- Improved spellchecking capabilities
- Improved integration with Apache Lucene
- New autosuggest component
- Distributed support for more components
- JSON document indexing and CSV response format
- Apache UIMA integration for metadata extraction
- Many other Bugfixes, improvements and optimizations

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

### 3.8 25 June 2010 - Solr 1.4.1 Released

Solr 1.4.1 has been released and is now available for public [download](#)! Solr 1.4.1 is a bug fix release for Solr 1.4 that includes many Solr bug fixes as well as Lucene bug fixes from Lucene 2.9.3.

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

### 3.9 7 May 2010 - Apache Lucene Eurocon 2010 Coming to Prague May 18-21

On May 18th to the 21st Prague will play host to the first ever dedicated Lucene and Solr User Conference in Europe: [Apache Lucene Eurocon 2010](#). This is a not-for-profit conference presented by Lucid Imagination, with net proceeds being donated to The Apache Software Foundation. [Registration is now open](#). Schedule highlights include:

- [Two days of in depth training classes](#):
  - Solr Application Development Workshop - Erik Hatcher
  - Lucene Bootcamp - Grant Ingersoll
- [Four general sessions](#):



- The Search Revolution: How Lucene & Solr Are Changing The World - Eric Gries
- From Publisher To Platform: How The Guardian Used Content, Search, and Open Source To Build a Powerful New Business Model - Stephen Dunn
- Software Disruption: How Using Open Source, Search, Big Data and Cloud technology are Disrupting IT - Zack Urlocker
- Solr 1.5 and Beyond - Yonik Seeley
- [24 technical sessions, spanning two days, divided into two tracks](#)
- [A Thursday night MeetUp](#)
- [An event at the Czech Beer Festival](#)

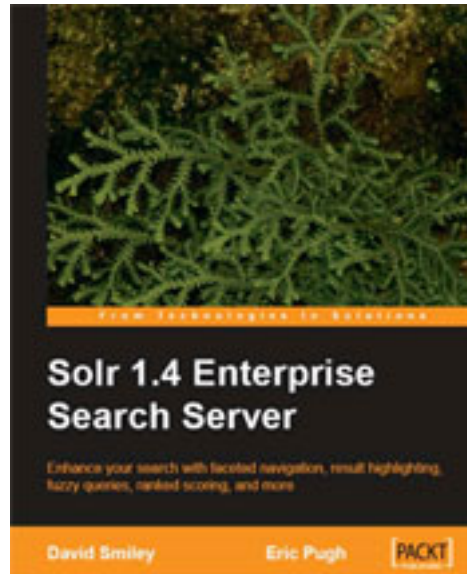
### 3.10 10 November 2009 - Solr 1.4 Released

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.

### 3.11 20 August 2009 - Solr's first book is published!



David Smiley and Eric Pugh are proud to introduce the first book on Solr, "[Solr 1.4 Enterprise Search Server](#)" from Packt Publishing.

This book is a comprehensive reference guide for nearly every feature Solr has to offer. It serves the reader right from initiation to development to deployment. It also comes with complete running examples to demonstrate its use and show how to integrate it with other languages and frameworks.

To keep this interesting and realistic, it uses a large open source set of metadata about artists, releases, and tracks courtesy of the MusicBrainz.org project. Using this data as a testing ground for Solr, you will learn how to import this data in various ways from CSV to XML to database access. You will then learn how to search this data in a myriad of ways, including Solr's rich query syntax, "boosting" match scores based on record data and other means, about searching across multiple fields with different boosts, getting facets on the results, auto-complete user queries, spell-correcting searches, highlighting queried text in search results, and so on.

After this thorough tour, you'll see working examples of integrating a variety of technologies with Solr such as Java, JavaScript, Drupal, Ruby, PHP, and Python.

Finally, this book covers various deployment considerations to include indexing strategies and performance-oriented configuration that will enable you to scale Solr to meet the needs of a high-volume site.

### 3.12 18 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

### 3.13 09 February 2009 - Lucene at ApacheCon Europe 2009 in Amsterdam

Lucene will be extremely well represented at [ApacheCon US 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

### 3.14 19 December 2008 - Solr Logo Contest Results

Many great logos were submitted, but only one could be chosen. Congratulations Michiel, the creator of the winning logo that is proudly displayed at the top of this page.

### 3.15 03 October 2008 - Solr Logo Contest

By popular demand, Solr is holding a contest to pick a new Solr logo. Details about how to submit an entry can be found [on the wiki](#). The Deadline for submissions is November 20th, 2008 @ 11:59PM GMT.

### 3.16 15 September 2008 - Solr 1.3.0 Available

Solr 1.3.0 is available for public download. This version contains many enhancements and 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](#).

### 3.17 28 August 2008 - Lucene/Solr at ApacheCon New Orleans



Lucene will be extremely well represented at [ApacheCon US 2008](#) in New Orleans this November 3-7, 2008:

- [Lucene Boot Camp](#) - A two day training session, November 3rd & 4th
- [Solr Boot Camp](#) - A one day training session, November 4th
- An [entire day of Lucene sessions](#) on November 5th

### 3.18 03 September 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.

### 3.19 06 June 2007: Release 1.2 available

This is the first release since Solr graduated from the Incubator, bringing many new features, including CSV/delimited-text data loading, time based autocommit, faster faceting, negative filters, a spell-check handler, sounds-like word filters, regex text filters, and more flexible plugins.

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

### 3.20 17 January 2007: Solr graduates from Incubator

Solr has graduated from the Apache Incubator, and is now a sub-project of Lucene.

### 3.21 22 December 2006: Release 1.1.0 available

This is the first release since Solr joined the Incubator, and brings many new features and performance optimizations including highlighting, faceted search, and JSON/Python/Ruby response formats.

### **3.22 15 August 2006: Solr at ApacheCon US**

Chris Hostetter will be presenting "[Faceted Searching With Apache Solr](#)" at ApacheCon US 2006, on October 13th at 4:30pm. See the [ApacheCon](#) website for more details.

### **3.23 21 April 2006: Solr at ApacheCon**

Yonik Seeley will be presenting "**Apache Solr, a Full-Text Search Server based on Lucene**" at ApacheCon Europe 2006, on June 29th at 5:30pm. See the [ApacheCon](#) website for more details.

### **3.24 21 February 2006: nightly builds**

Solr now has nightly builds. This automatically creates a [downloadable version of Solr every night](#). All unit tests must pass, or a message is sent to the developers mailing list and no new version is created. This also updates the [javadoc](#).

### **3.25 17 January 2006: Solr Joins Apache Incubator**

Solr, a search server based on Lucene, has been accepted into the Apache Incubator. Solr was originally developed by CNET Networks, and is widely used within CNET to provide high relevancy search and faceted browsing capabilities.