Welcome to Lucene!

Table of contents

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 algorithmatic 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. 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

2.2. 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.

2.3. 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.

2.4. 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

2.5. 19 January 2009 - PyLucene joins the Lucene TLP

PyLucene, the Python based port of Lucene is now an official Lucene subproject.

2.6. 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.

2.7. 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.