


Lucene

Lucene

	
Developer(s)	Apache Software Foundation
Stable release	4.3.0 / May 6, 2013
Development status	Active
Written in	Java
Operating system	Cross-platform
Type	Search and index
License	Apache License 2.0
Website	lucene.apache.org ^[1]

Apache Lucene is a free/open source information retrieval software library, originally created in Java by Doug Cutting. It is supported by the Apache Software Foundation and is released under the Apache Software License.

Lucene has been ported to other programming languages including Delphi, Perl, C#, C++, Python, Ruby, and PHP.^[2]

History

Doug Cutting originally wrote Lucene in 1999.^[3] It was initially available for download from its home at the SourceForge web site. It joined the Apache Software Foundation's Jakarta family of open-source Java products in September 2001 and became its own top-level Apache project in February 2005. Until recently, Wikipedia:Manual of Style/Dates and numbers#Chronological items it included a number of sub-projects, such as Lucene.NET, Mahout, Solr and Nutch. Solr has merged into the Lucene project itself and Mahout, Nutch, and Tika have moved to become independent top-level projects.

Features and common use

While suitable for any application which requires full text indexing and searching capability, Lucene has been widely recognized^{[4][5]} for its utility in the implementation of Internet search engines and local, single-site searching.

At the core of Lucene's logical architecture is the idea of a **document** containing **fields** of text. This flexibility allows Lucene's API to be independent of the file format. Text from PDFs, HTML, Microsoft Word, and OpenDocument documents, as well as many others (except images), can all be indexed as long as their textual information can be extracted.^[6]

Lucene-based projects

Lucene itself is just an indexing and search library and does not contain crawling and HTML parsing functionality. However, several projects extend Lucene's capability:

- Apache Nutch — provides web crawling and HTML parsing
- Apache Solr — an enterprise search server
- ElasticSearch — an enterprise search server
- Compass — a Java Search Engine Framework
- DocFetcher — a multiplatform desktop search application

Lucene.NET

Lucene.NET is a port of Lucene written in C# and targeted at .NET Framework users. There are currently two variations of the software, differing in Generics support and a few bug fixes. The version that supports Generics has a lowercase "g" appended to the version name to distinguish between the two.^[citation needed]

Regardless of generics support, all versions of Lucene.NET (with or without the "g") should be able to read and write index files from the corresponding version of Java Lucene with no issues.^[citation needed]

Users

For a list of companies that use Lucene (rather than extend), see Lucene's "Powered By" page.^[7] As an example, Twitter is using Lucene for its real time search.^[8]

References

- [1] <http://lucene.apache.org>
- [2] Lucene implementations (<http://wiki.apache.org/lucene-java/LuceneImplementations>)
- [5] GNU/Linux Semantic Storage System (http://www.glscube.org/downloads/glscube_design.pdf)
- [7] PoweredBy (<http://wiki.apache.org/lucene-java/PoweredBy>)
- [8] Twitter uses Lucene (<http://techcrunch.com/2010/10/06/new-twitter-search/>)

Bibliography

- Gospodnetic, Otis; Erik Hatcher, Michael McCandless (28 June 2009). *Lucene in Action* (2nd ed.). Manning Publications. p. 475. ISBN 1-9339-8817-7.
- Gospodnetic, Otis; Erik Hatcher (1 December 2004). *Lucene in Action* (1st ed.). Manning Publications. p. 456. ISBN 978-1-9323-9428-3.

External links

- Official website (<http://lucene.apache.org/>)
- Lucene.NET (<http://lucenenet.apache.org/>)
- List of Lucene Ports (or Implementations) in Other Languages on the Apache wiki (<http://wiki.apache.org/lucene-java/LuceneImplementations>)
- Introductory article with Java code for search (<http://web.archive.org/web/20060715234923/schmidt.devlib.org/software/lucene-wikipedia.html>) (This page is on Internet Archive) on Wikipedia data (<http://download.wikimedia.org/wikipedia/>)
- An article with explanation and example implementation of Lucene.Net (http://www.codeminima.com/CSharp:_A_simple_Lucene.Net_indexing_and_search_class)

Article Sources and Contributors

Lucene *Source:* <http://en.wikipedia.org/w/index.php?oldid=553776071> *Contributors:* 900mill, Andy Dingley, AndyHedges, Anna Frodesiak, Apokrif, Arkham, AxelBoldt, Bawolff, Beland, Blago, Bonadea, Brew57, CVertex, Cej10, Chris the speller, Cobaltcigs, DStoykov, Dank, Darthbear, David Gerard, David Spencer, Davidbspalding, DeansFA, Dekisugi, Den fjättrade ankan, Dhiraj1984, Dnaber, Duffbeerforme, Eleassar777, Ellmist, Elmook, Emmanuel.keller, Frap, FuFoFuEd, Gaius Cornelius, Gary King, Gnewton, Gomeztoo, GrantIngersoll, Greggrestonva, Gurch, Guyjohnston, Gvasil, Gwern, Hao2lian, Hervegirod, Hibharani, Hu12, Igbrown, Ipsjolly, J04n, JLaTondre, Javasoze, Jeffiq, Jhertel, Jkonline, Jnareb, Jonnylamb, Joos23, Joshisachin79, KTC, Kesla, Khalid hassani, Kimchy77, K14m-AWB, Kocio, Korath, LAS1180, Latchkey, Leonstar, LittleBenW, Logixoul, Looptek, Lunacrescens, Lunarmys, Luís Felipe Braga, Machadofisher, MaikSchreiber, MartinWertheimer, Mcvos, Mhenry07, Mikemccand, Mion, Momma's Little Helper, Mongol, Mortense, NFOlson, Najeeb1010, Nyco, OlEnglish, Olekva, Open2universe, Pamri, Paul.daniell, Peatar, Phantom25, Pinethicket, Pmauriciocosta, Prasadgupte, Pushpinder12, R'n'B, RHaworth, Ray3055, RedWolf, Rich Farmbrough, Rkthorne, Rodrigoluk, Ronklein, Ronnie6657, Rror, Runarb, SF007, Sam Hocevar, Sfermigier, Shlomif, Shodanium, Simonwillnauer, Slau1364, Smb1001, SouthLake, Stephenb, Svincentc, Tdunning, Ted nw, Tejas14883, Thetaphi, Thorwald, Thumperward, Tobias382, Torc2, Tsiaojian lee, Utcursch, Vinayaraj, Vineet Sinha MIT, Wickethewok, Winterst, Wolfman, Yangeffrotin, Yonkeltron, Zbxgscqf, ZeroOne, Zigger, Zodeh, 233 anonymous edits

Image Sources, Licenses and Contributors

Image:Lucene logo green 300.png *Source:* http://en.wikipedia.org/w/index.php?title=File:Lucene_logo_green_300.png *License:* GNU Free Documentation License *Contributors:* Beao, Bennylin, Jonkerz, Yangeffrotin

License

Creative Commons Attribution-Share Alike 3.0 Unported
[//creativecommons.org/licenses/by-sa/3.0/](http://creativecommons.org/licenses/by-sa/3.0/)