Welcome to Nutch!

Table of contents

1. News

1.1. 06 June 2010 - Apache Nutch 1.1 Released

The 1.1 release of Nutch is now available. This release includes several major upgrades of existing libraries (Hadoop, Solr, Tika, etc.) on which Nutch depends. Various bug fixes, and speedups (e.g., to Fetcher2) have also been included. See list of changes made in this version. The release is available here.

1.2. 21 April 2010 - Apache Nutch graduates to TLP

Passed by unanimous approval of the Apache Board, Nutch graduated to TLP status. We are in the process of updating the website, and moving things around, so if you notice anything out of place, please let us know.

1.3. 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

1.4. 23 March 2009 - Apache Nutch 1.0 Released

The 1.0 release of Nutch is now available. This release includes several major feature improvements such as new indexing framework, new scoring framework, Apache Solr integration just to mention a few. See [list of changes](#) made in this version. The release is available [here](#).

1.5. 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

1.6. 2 April 2007: Nutch 0.9 Released

The 0.9 release of Nutch is now available. This is the second release of Nutch based entirely on the underlying Hadoop platform. This release includes several critical bug fixes, as well as key speedups described in more detail at [Sami Siren's blog](#). See [list of changes](#) made in this version. The release is available [here](#).

1.7. 24 September 2006: Nutch 0.8.1 Released

The 0.8.1 release of Nutch is now available. This is a maintenance release to 0.8 branch fixing many serous bugs found in version 0.8. See list of changes made in this version. The release is available here.

1.8. 25 July 2006: Nutch 0.8 Released

The 0.8 release of Nutch is now available. This is the first release of Nutch based on hadoop architecure. See CHANGES.txt for list of changes made in this version. The release is available here.

1.9. 31 March 2006: Nutch 0.7.2 Released

The 0.7.2 release of Nutch is now available. This is a bug fix release for 0.7 branch. See CHANGES.txt for details. The release is available here.

1.10. 1 October 2005: Nutch 0.7.1 Released

The 0.7.1 release of Nutch is now available. This is a bug fix release. See CHANGES.txt for details. The release is available here.

1.11. 17 August 2005: Nutch 0.7 Released

This is the first Nutch release as an Apache Lucene sub-project. See CHANGES.txt for details. The release is available here.

1.12. June 2005: Nutch graduates from Incubator

Nutch has now graduated from the Apache incubator, and is now a Subproject of Lucene.

1.13. January 2005: Nutch Joins Apache Incubator

Nutch is a two-year-old open source project, previously hosted at Sourceforge and backed by its own non-profit organization. The non-profit was founded in order to assign copyright, so that we could retain the right to change the license. We have now determined that the Apache license is the appropriate license for Nutch and no longer require the overhead of an independent non-profit organization. Nutch's board of directors and its developers were both polled and supported the move to the Apache foundation.

1.14. September 2004: Creative Commons launches Nutch-based Search

Creative Commons unveiled a beta version of its search engine, which scours the web for text, images, audio, and video free to re-use on certain terms a search refinement offered by no other company or organization.

See the [Creative Commons Press Release](#) for more details.

1.15. September 2004: Oregon State University switches to Nutch

Oregon State University is converting its searching infrastructure from Googletm to the open source project Nutch. The effort to replace the Googletm will realize significant cost savings for Oregon State University, while promoting both the Nutch Search Engine and transparency in search engine use and management.

For more details see the announcement by OSU's [Open Source Lab](#).