

Maintaining Open Source Software 20 Years from Today

Marcus Eagan



\$ whoami and \$ whyicare

Lucene History

Maintain

Improve

Extend



Doug Cutting



Doug's Search Lineage

parc[®]
A Xerox Company

parc[®]
A Xerox Company



excite



An Object-Oriented Architecture for Text Retrieval

**Doug Cutting, Jan Pedersen, and
Per-Kristian Halvorsen**

Xerox Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, California

Abstract

For almost all aspects of information access systems it is still the case that their optimal composition and functionality is hotly debated. Moreover, different application scenarios put different demands on individual components. It is therefore of the essence to be able to quickly build systems that permit exploration of different designs and implementation strategies. This paper presents a software implementation architecture for text retrieval systems that facilitates (a) functional modularization (b) mix-and-match combination of module implementations and (c) definition of inter-module protocols. We show how an object-oriented approach easily accommodates this type of architecture. The design principles are exemplified by code examples in Common Lisp. Taken together these code examples constitute an operational retrieval system. The design principles and protocols implemented have also been instantiated in a large scale retrieval prototype in our research laboratory.



Early Days

Lucene Joins Apache Jakarta

2001

Apache Lucene <-> Solr Merger

2010

Apache Lucene files for Divorce

2021



Apache Lucene

- Donated by legendary search engineer Doug Cutting to Apache Software Foundation in 2001
- Regular contributors and users from all over the world, six continents.
- The backbone of several projects and search engines you use every day
- Packages a lot of the cutting edge research from information retrieval, data science, natural language processing, and computing in general
 - The project has surfaced several bugs in the JVM!



Maintenance



**CLOSE ONE TICKET, TWO MORE
SHALL TAKE ITS PLACE.**



HAIL JIRA



Code Clarity

```
SOLR-14702: Remove oppressive language (part1) #1711
Changes from all commits - File filter - Beta Conversations - Jump to - Viewed

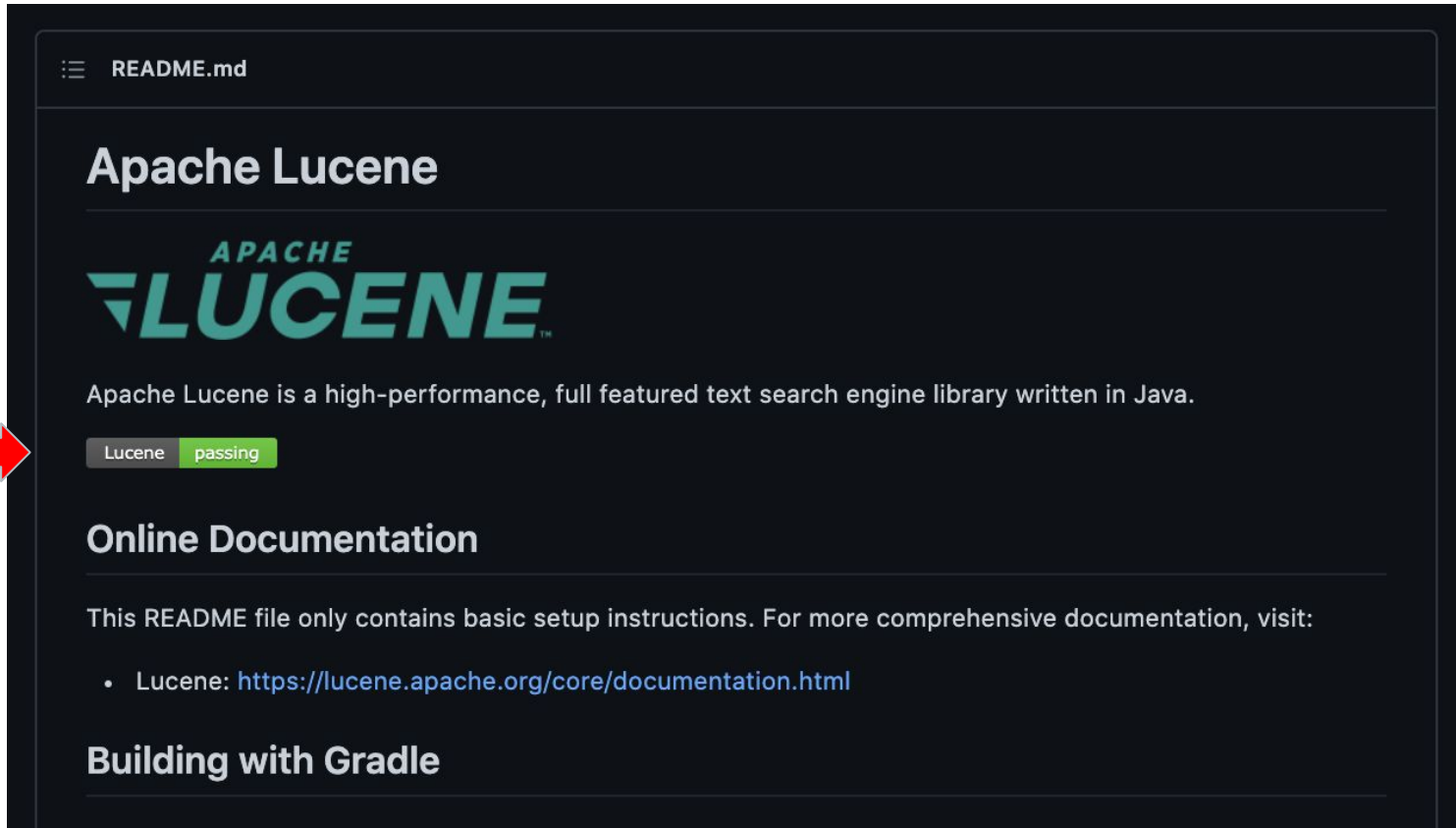
solr/contrib/prometheus-exporter/conf/solr-exporter-config.xml

@@ -985,7 +985,7 @@
985     (if $parent_key_item_len == 5 then $parent_key_items[3] else "" end) as $shard |
986     (if $parent_key_item_len == 5 then $parent_key_items[4] else "" end) as $replica |
987     (if $parent_key_item_len == 5 then ($collection + "-" + $shard + "-" + $replica) else $score end) as $score |
988 -     $parent.value | to_entries | .[] | select(.key == "REPLICATION./replication.isMaster") as $subject |
989     $subject.key | split(".") [0] as $category |
990     $subject.key | split(".") [1] as $handler |
991     (if $subject.value == true then 1.0 else 0.0 end) as $value |

@@ -1018,13 +1018,13 @@
1018     (if $parent_key_item_len == 5 then $parent_key_items[3] else "" end) as $shard |
1019     (if $parent_key_item_len == 5 then $parent_key_items[4] else "" end) as $replica |
1020     (if $parent_key_item_len == 5 then ($collection + "-" + $shard + "-" + $replica) else $score end) as $score |
1021 -     $parent.value | to_entries | .[] | select(.key == "REPLICATION./replication.isSlave") as $subject |
1022     $subject.key | split(".") [0] as $category |
1023     $subject.key | split(".") [1] as $handler |
1024     (if $subject.value == true then 1.0 else 0.0 end) as $value |
1025     if $parent_key_item_len == 3 then
1026     {
1027 -     name: "solr_metrics_core_replication_slave",
1028     type: "GAUGE",
1029     help: "See following URL: https://lucene.apache.org/solr/guide/metrics-reporting.html",
1030     label_names: ["category", "handler", "core"],
1031     }
1032     }
1033     }
1034     else
1035     {
1036 -     name: "solr_metrics_core_replication_slave",
1037     type: "GAUGE",
1038     help: "See following URL: https://lucene.apache.org/solr/guide/metrics-reporting.html",
1039     label_names: ["category", "handler", "core", "collection", "shard", "replica"],
1040     }
1041     }
1042     }


@@ -1021,7 +1021,7 @@
1021     $parent.value | to_entries | .[] | select(.key == "REPLICATION./replication.isFollower") as $subject |
1022     $subject.key | split(".") [0] as $category |
1023     $subject.key | split(".") [1] as $handler |
1024     (if $subject.value == true then 1.0 else 0.0 end) as $value |
1025     if $parent_key_item_len == 3 then
1026     {
1027 +     name: "solr_metrics_core_replication_follower",
1028     type: "GAUGE",
1029     help: "See following URL: https://lucene.apache.org/solr/guide/metrics-reporting.html",
1030     label_names: ["category", "handler", "core"],
1031     }
1032     }
1033     }
1034     else
1035     {
1036 +     name: "solr_metrics_core_replication_follower",
1037     type: "GAUGE",
1038     help: "See following URL: https://lucene.apache.org/solr/guide/metrics-reporting.html",
1039     label_names: ["category", "handler", "core", "collection", "shard", "replica"],
1040     }
1041     }
1042     }
```

Project Awareness



☰ README.md

Apache Lucene



Apache Lucene is a high-performance, full featured text search engine library written in Java.

Lucene passing

Online Documentation

This README file only contains basic setup instructions. For more comprehensive documentation, visit:

- Lucene: <https://lucene.apache.org/core/documentation.html>

Building with Gradle



Project Health

Analyzer performance over time in Lucene trunk



Improvement

“The principal goal of the engineer is to simplify.” - Erik Hatcher, the author of Lucene in Action, Co-Founder of Lucidworks, and brother from another mother



Don't Alert Security Team

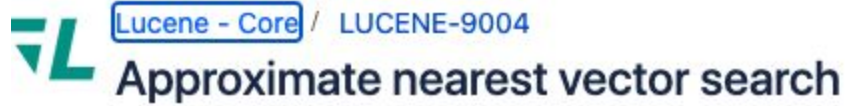
SOLR-15453 - CSP Fix Last Week

LUCENE-9420 - Vulnerable Dependency



SOLR-13649 - Block Unknown by Default



Vector Search



Details

Type:	 New Feature
Priority:	 Major
Affects Version/s:	None
Component/s:	None
Labels:	None
Lucene Fields:	New



The Next Big Thing

Cue Mark Miller




Ref Impl



Solr / SOLR-14788

Solr: The Next Big Thing

Details

Type: Task
Priority:  Critical
Affects Version/s: None
Component/s: None
Labels: None

Status: **OPEN**
Resolution: Unresolved
Fix Version/s: None

Description

















***The Policeman is NOW OFF duty!**

When The Policeman is on duty, sit back, relax, and have some fun. Try to make some progress. Don't stress too much about the impact of your changes or maintaining stability and performance and correctness so much. Until the end of phase 1, I've got your back. I have a variety of tools and contraptions I have been building over the years and I will continue training them on this branch. I will review your changes and peer out across the land and course correct where needed. As Mike D will be thinking, "Sounds like a bottleneck Mark." And indeed it will be to some extent. Which is why once stage one is completed, I will flip The Policeman to off duty. When off duty, I'm always occasionally *down for some vigilante justice, but I won't be walking the beat, all that stuff about sit back and relax goes out the window.



Ref Impl

Sub-Tasks

- | | | | |
|-----|--|--|----------------------|
| 1. | Remove added ignores to tests that still have them |  OPEN | Mark Robert Miller |
| 2. |  Bring back shard whitelisting feature |  RESOLVED | Noble Paul |
| 3. | Handle security in the node forwarding / proxy / remoteQuery feature that HttpSolrCall has |  OPEN | Noble Paul |
| 4. |  Allow Rule Based Replica Assignment Policy |  RESOLVED | Unassigned |
| 5. |  Wrap up async issues and tests. |  RESOLVED | Mark Robert Miller |
| 6. | Synchronizing reference and master branches |  OPEN | Ishan Chattopadhyaya |
| 7. | Address Alias impl. |  OPEN | Unassigned |
| 8. |  Address Tolerant search feature and tests. |  RESOLVED | Mark Robert Miller |
| 9. | Demonstrate how Solr 8, master, or any version previous Solr version before pales next to the reference branch. |  OPEN | Unassigned |
| 10. | Nightly test run. |  OPEN | Unassigned |



SPECIAL THANK YOUs

Starting with Nina Mueller, Plain Shwarz, Yonik Seeley, and Elastic (especially the homies Uri and Simon)



ERIK HATCHER and SHALIN MANGAR

MIKE's SOKOLOV and McCANDLESS

TIM POTTER AND ERICK ERICKSON

JAN HØYDAHL and ROBERT MUIR

MIKE DROB

KEVIN RISDEN

NOBLE PAUL

TOMAS E. F. L.

ISHAN CHATTOPADHYAY



JAREN GLOVER

DAVID SMILEY

JOEL BERNSTEIN

HOUSTON PUTMAN

ALEA ABED

WILL HAYES

VIVEK SRIRAM

KORD & GRANDMA JOANN





@marcusforpeace

I tweet about open source, startups I invest in/don't
invest in

DMs open