


How to search

...better!

Alessandro Melchiori // @amelchiori

Unstructured search



GitHub Explore Features Enterprise Blog [Sign up](#) [Sign in](#)

Search [Search](#)

Repositories 317

- [Code](#) 17,981
- [Issues](#) 2,008
- [Users](#) 2

Languages

- Java** 167
- Ruby 139
- JavaScript 117
- Python 69
- PHP 49
- Shell 40
- Puppet 38
- Perl 16
- Scala 13
- C#

We've found 317 repository results Sort: **Best match**

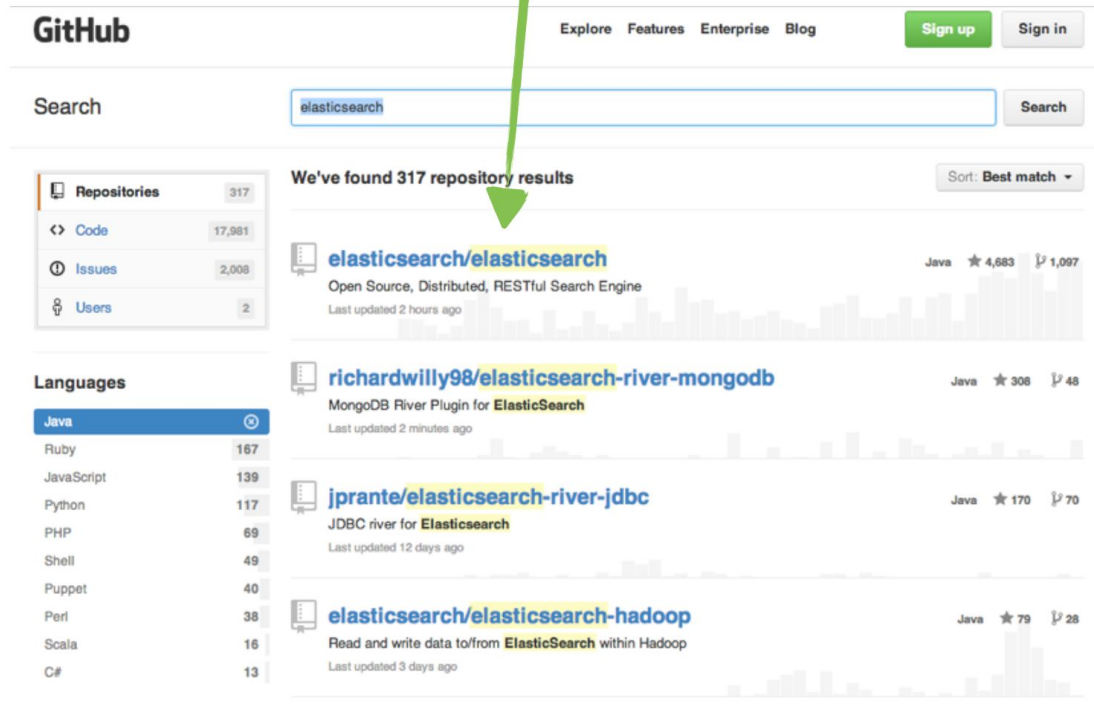
elasticsearch/elasticsearch Java ★ 4,683 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb Java ★ 308 48
MongoDB River Plugin for **ElasticSearch**
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc Java ★ 170 70
JDBC river for **Elasticsearch**
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop Java ★ 79 28
Read and write data to/from **ElasticSearch** within Hadoop
Last updated 3 days ago

Enrichment



The screenshot shows the GitHub search interface. At the top, the GitHub logo is on the left, and navigation links for 'Explore', 'Features', 'Enterprise', and 'Blog' are on the right. Below the navigation bar is a search bar containing the text 'elasticsearch' and a 'Search' button. To the left of the search results is a sidebar with filters for 'Repositories' (317), 'Code' (17,981), 'Issues' (2,008), and 'Users' (2). Below the filters is a 'Languages' section with a list of programming languages and their respective repository counts. The main search results area displays 'We've found 317 repository results' and a 'Sort: Best match' dropdown. The top three results are listed, each with a repository icon, the repository name, a brief description, and a bar chart showing activity over time. A green arrow points from the word 'Enrichment' at the top of the page to the first result, 'elasticsearch/elasticsearch'.

GitHub Explore Features Enterprise Blog [Sign up](#) [Sign in](#)

Search [Search](#)

Repositories 317

- [Code](#) 17,981
- [Issues](#) 2,008
- [Users](#) 2

Languages

- Java** 317
- Ruby 167
- JavaScript 139
- Python 117
- PHP 69
- Shell 49
- Puppet 40
- Perl 38
- Scala 16
- C# 13

We've found 317 repository results [Sort: Best match](#)

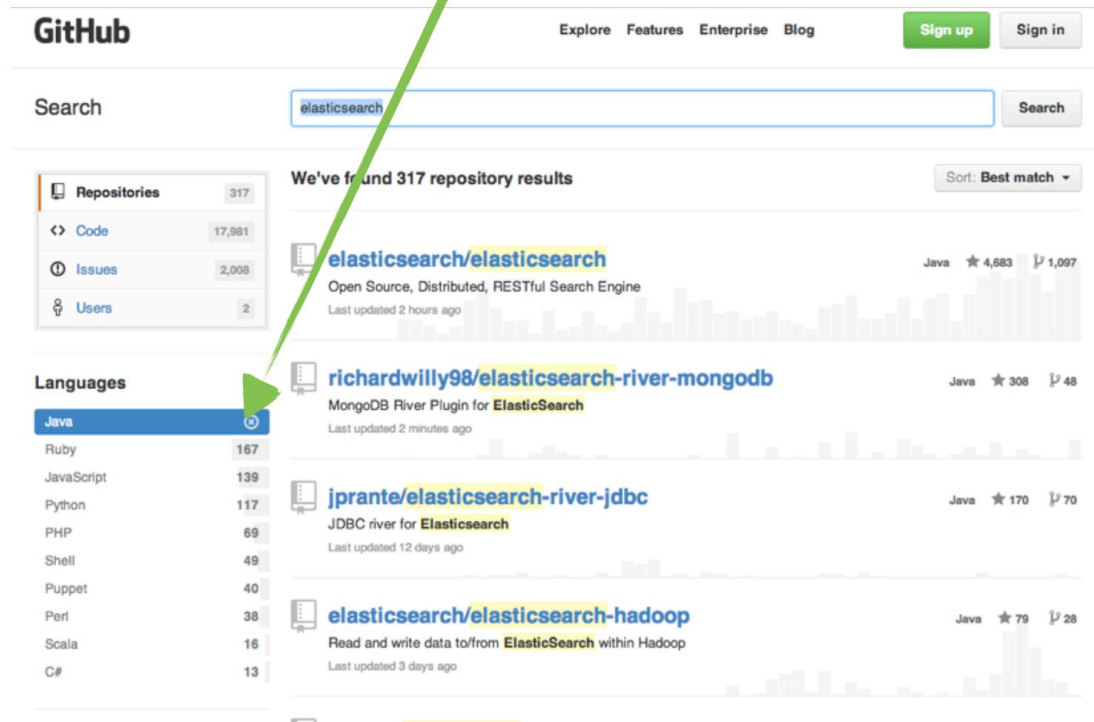
elasticsearch/elasticsearch Java ★ 4,683 📄 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb Java ★ 308 📄 48
MongoDB River Plugin for **ElasticSearch**
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc Java ★ 170 📄 70
JDBC river for **Elasticsearch**
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop Java ★ 79 📄 28
Read and write data to/from **ElasticSearch** within Hadoop
Last updated 3 days ago

Aggregation



The screenshot shows the GitHub search interface. At the top, the GitHub logo is on the left, and navigation links for 'Explore', 'Features', 'Enterprise', and 'Blog' are on the right, along with 'Sign up' and 'Sign in' buttons. Below the header is a search bar containing 'elasticsearch' and a 'Search' button. To the left of the search results is a sidebar with filters. The 'Repositories' filter shows 317 results. Below it are links for 'Code' (17,981), 'Issues' (2,006), and 'Users' (2). The 'Languages' filter is expanded, showing a list of languages with 'Java' selected and highlighted in blue. A green arrow points from the word 'Aggregation' at the top of the slide to the 'Java' language filter. The main search results area displays 'We've found 317 repository results' and a 'Sort: Best match' dropdown. The first three results are listed, each with a repository name, description, last update time, and a bar chart showing activity over time.

Repository	Description	Last updated	Stars	Forks
elasticsearch/elasticsearch	Open Source, Distributed, RESTful Search Engine	Last updated 2 hours ago	4,583	1,097
richardwilly98/elasticsearch-river-mongodb	MongoDB River Plugin for ElasticSearch	Last updated 2 minutes ago	306	48
jprante/elasticsearch-river-jdbc	JDBC river for Elasticsearch	Last updated 12 days ago	170	70
elasticsearch/elasticsearch-hadoop	Read and write data to/from ElasticSearch within Hadoop	Last updated 3 days ago	79	28

Suggestions



GitHub

This repository ▾ debian

Sign up Sign in

Star 4,683 Fork 1,097

New Issue

1 2 3 ... 19

Labels

- Lucene 4.5 Upgrade
- breaking
- bug
- enhancement
- feature
- non-issue

elasticsearch/elasticsearch#1726 debian package violates naming convention

elasticsearch/elasticsearch#3571 debian package init-script: start-stop-daemon ne

elasticsearch/elasticsearch#1681 Debian pkg

elasticsearch/elasticsearch#3286 There is no official debian/ubuntu repository

elasticsearch/elasticsearch#3500 Elasticsearch should include debian's standard j

elasticsearch/elasticsearch#1526 Moving debian package to maven

Search elasticsearch/elasticsearch for 'debian'

Search GitHub for 'debian'

1 Opened by s1monw 14 hours ago

11

10

9

1

NoShardAvailableActionException in ES 0.90.3 on startup
Opened by richardwilly98 a day ago

Feature Request: Don't reindex the document when updating non-indexed fields
Opened by ddorian 2 days ago 4 comments

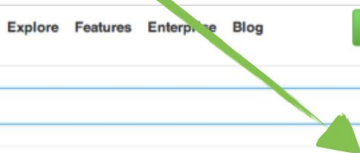
#3702

#3701

#3700

#3696

Sorting



GitHub Explore Features Enterprise Blog [Sign up](#) [Sign in](#)

Search [Search](#)

We've found 317 repository results [Sort: Best match ▾](#)

Repositories 317

- [Code](#) 17,981
- [Issues](#) 2,008
- [Users](#) 2

Languages

- Java** 317
- Ruby 167
- JavaScript 139
- Python 117
- PHP 69
- Shell 49
- Puppet 40
- Perl 38
- Scala 16
- C# 13

elasticsearch/elasticsearch Java ★ 4,683 1,097
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago

richardwilly98/elasticsearch-river-mongodb Java ★ 308 48
MongoDB River Plugin for **ElasticSearch**
Last updated 2 minutes ago

jprante/elasticsearch-river-jdbc Java ★ 170 70
JDBC river for **Elasticsearch**
Last updated 12 days ago

elasticsearch/elasticsearch-hadoop Java ★ 79 28
Read and write data to/from **ElasticSearch** within Hadoop
Last updated 3 days ago

Pagination

The screenshot shows the GitHub search interface. At the top, the GitHub logo is on the left, and navigation links (Explore, Features, Enterprise, Blog) and buttons (Sign up, Sign in) are on the right. Below the header is a search bar containing 'elasticsearch' and a 'Search' button. To the left of the search results is a sidebar with filters: Repositories (317), Code (17,981), Issues (2,008), and Users (2). Below the filters is a 'Languages' section with a list of languages including Java, Ruby, JavaS, Python, PHP, Shell, Puppet, Perl, Scala, and C#. The main search results area shows 'We've found 317 repository results' with a 'Sort: Best match' dropdown. The first result is 'elasticsearch/elasticsearch', an Open Source, Distributed, RESTful Search Engine, last updated 2 hours ago, with 4,683 stars and 1,097 forks. The second result is 'spinscale/elasticsearch-suggest-plugin', a Plugin for elasticsearch which uses the lucene FSTSuggester, last updated 4 days ago, with 103 stars and 23 forks. A green arrow points from the word 'Pagination' at the top of the slide to the pagination controls at the bottom of the second result. The pagination controls show a sequence of numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9, ..., 31, 32. The number 1 is highlighted, indicating the current page. To the right of the pagination controls is a link that says 'How are these search results? Tell us!'.

GitHub

Explore Features Enterprise Blog Sign up Sign in

Search elasticsearch Search

We've found 317 repository results Sort: Best match

Repositories 317
Code 17,981
Issues 2,008
Users 2

Languages
Java
Ruby
JavaS
Python
PHP
Shell
Puppet
Perl
Scala
C#

elasticsearch/elasticsearch
Open Source, Distributed, RESTful Search Engine
Last updated 2 hours ago
Java ★ 4,683 1,097

spinscale/elasticsearch-suggest-plugin
Plugin for elasticsearch which uses the lucene FSTSuggester
Last updated 4 days ago
Java ★ 103 23

1 2 3 4 5 6 7 8 9 ... 31 32

How are these search results? Tell us!

Welcome Azure Search!

Azure Search is a search-as-a-service cloud solution that gives developers APIs and tools for adding a rich search experience over private, heterogenous content in web, mobile, and enterprise applications.

DEMO

[index and mapping]

Inverted index

"The quick brown fox jumped over the lazy dog"

"Quick brown foxes leap over lazy dogs in summer"

Inverted index: **separate words / terms**

[The, quick, brown, fox, jumped, over, the, lazy, dog]

[Quick, brown, foxes, leap, over, lazy, dogs, in, summer]

Inverted index: **sort unique terms**

[The,brown,dog,fox,jumped,lazy,over,quick,the]

[Quick,brown,dogs,foxes,in,lazy,leap,over,summer]

Inverted index: **list docs containing terms**

term	doc-01	doc-02
Quick		X
The	X	
brown	X	X
dog	X	
dogs		X
fox	X	
...

analysis

=

tokenization + normalization

analyzer

=

tokenizer + token filters

Standard analyzer

`"The Quick Brown Fox jumped over the Lazy Dog!"`

Standard analyzer **standard tokenizer**

`"The, Quick, Brown, Fox, jumped, over, the, Lazy, Dog"`

Standard analyzer

lowercase filter

`"the, quick, brown, fox, jumped, over, the, lazy, dog"`

Standard analyzer **stopword filter**

`"quick,brown,fox,jumped,over,lazy,dog"`

English analyzer

`"The Quick Brown Fox jumped over the Lazy Dog!"`

English analyzer

standard tokenizer

`"The, Quick, Brown, Fox, jumped, over, the, Lazy, Dog"`

English analyzer

lowercase filter

`"the, quick, brown, fox, jumped, over, the, lazy, dog"`

English analyzer

english stemmer

`"the, quick, brown, fox, jumped, over, the, lazy, dog"`

English analyzer

english stemmer

`"the, quick, brown, fox, jump, over, the, lazy, dog"`

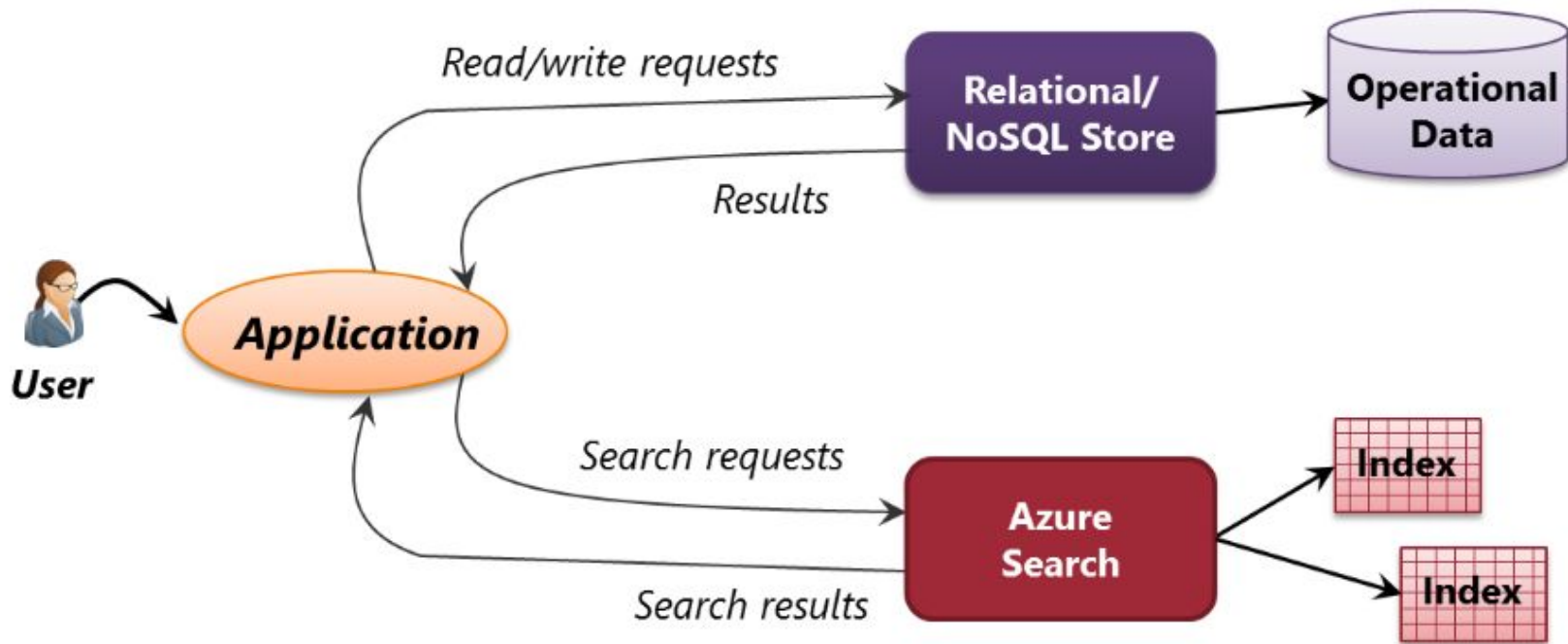
English analyzer

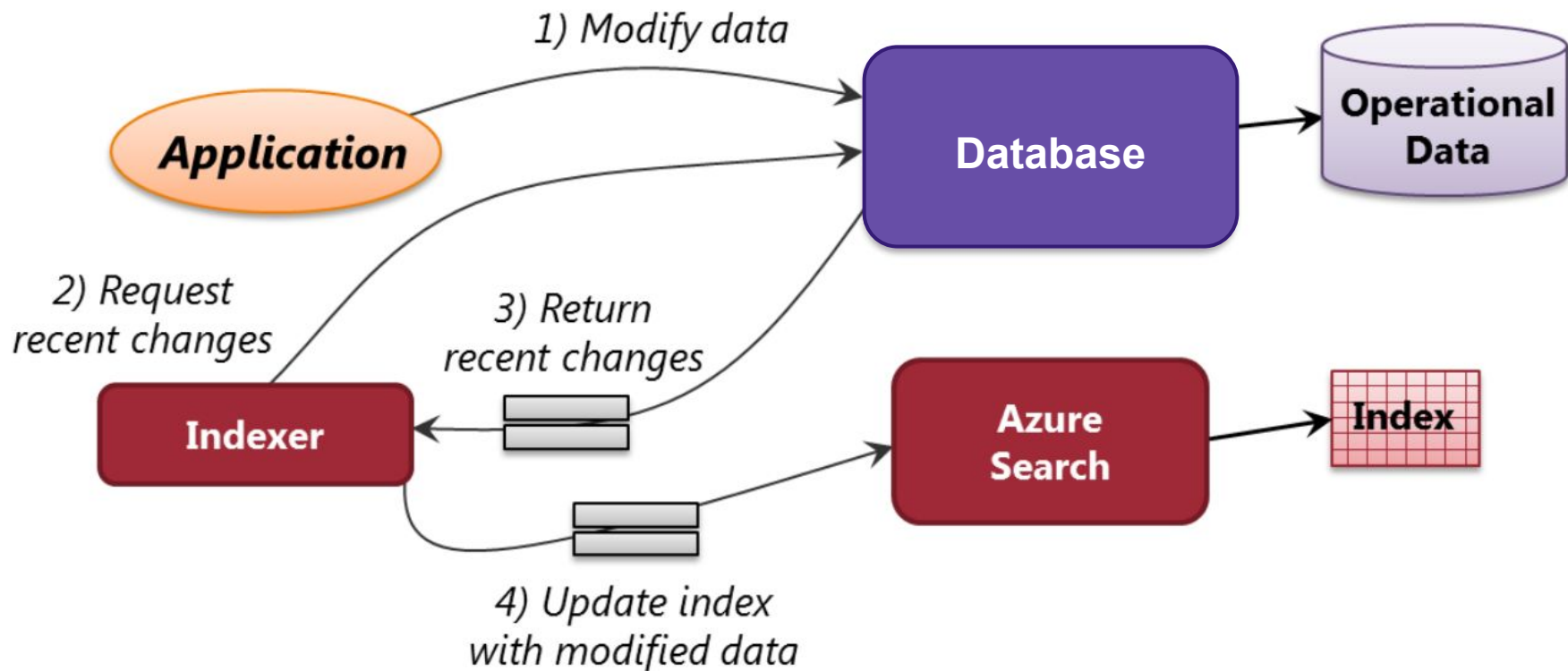
english stopwords

`"quick,brown,fox,jump,over,lazy,dog"`

DEMO

[advanced search]





DEMO

[manual and automatic indexer]



Alessandro Melchiori

Software developer @ CodicePlastico
alessandro@codiceplastico.com
[@amelchiori](https://twitter.com/amelchiori)