

# Lucene Query Syntax

Lucene has a custom query syntax for querying its indexes. Here are some query examples demonstrating the query syntax.

## Keyword matching

Search for word "foo" in the title field.

```
title:foo
```

Search for phrase "foo bar" in the title field.

```
title:"foo bar"
```

Search for phrase "foo bar" in the title field AND the phrase "quick fox" in the body field.

```
title:"foo bar" AND body:"quick fox"
```

Search for either the phrase "foo bar" in the title field AND the phrase "quick fox" in the body field, or the word "fox" in the title field.

```
(title:"foo bar" AND body:"quick fox") OR title:fox
```

Search for word "foo" and not "bar" in the title field.

```
title:foo -title:bar
```

## Wildcard matching

Search for any word that starts with "foo" in the title field.

```
title:foo*
```

Search for any word that starts with "foo" and ends with bar in the title field.

```
title:foo*bar
```

Note that Lucene doesn't support using a \* symbol as the first character of a search.

## Proximity matching

Lucene supports finding words are a within a specific distance away.

Search for "foo bar" within 4 words from each other.

```
"foo bar"~4
```

Note that for proximity searches, exact matches are proximity zero, and word transpositions (bar foo) are proximity 1.

A query such as **"foo bar"~10000000** is an interesting alternative to **foo AND bar**.

Whilst both queries are effectively equivalent with respect to the documents that are returned, the proximity query assigns a higher score to documents for which the terms **foo** and **bar** are closer together.

The trade-off, is that the proximity query is slower to perform and requires more CPU.

Solr DisMax and eDisMax query parsers can add phrase proximity matches to a user query.

## Range searches

Range Queries allow one to match documents whose field(s) values are between the lower and upper bound specified by the Range Query. Range Queries can be inclusive or exclusive of the upper and lower bounds. Sorting is done lexicographically.

```
mod_date:[20020101 TO 20030101]
```

Solr's built-in field types are very convenient for performing range queries on numbers without requiring padding.

## Boosts

Query-time boosts allow one to specify which terms/clauses are "more important". The higher the boost factor, the more relevant the term will be, and therefore the higher the corresponding document scores.

A typical boosting technique is assigning higher boosts to title matches than to body content matches:

```
(title:foo OR title:bar)^1.5 (body:foo OR body:bar)
```

You should carefully examine **explain** output to determine the appropriate boost weights.

The official docs for the query parser syntax are here: [http://lucene.apache.org/java/3\\_5\\_0/queryparsersyntax.html](http://lucene.apache.org/java/3_5_0/queryparsersyntax.html) ([http://lucene.apache.org/java/3\\_5\\_0/queryparsersyntax.html](http://lucene.apache.org/java/3_5_0/queryparsersyntax.html))

The query syntax has not changed significantly since Lucene 1.3 (it is now 3.5.0).

## Parsing Queries

Queries can be parsed by constructing a QueryParser object and invoking the parse() method.

```
String querystr = args.length > 0 ? args[0] : "lucene";
```

```
Query q = new QueryParser(Version.LUCENE_CURRENT, "title", analyzer).parse(querystr);
```

## Getting Started

Home ( <a href="http://www.lucenetutorial.com/index.html">http://www.lucenetutorial.com/index.html</a> )
Lucene in 5 Minutes ( <a href="http://www.lucenetutorial.com/lucene-in-5-minutes.html">http://www.lucenetutorial.com/lucene-in-5-minutes.html</a> )
Lucene NRT Example ( <a href="http://www.lucenetutorial.com/lucene-nrt-hello-world.html">http://www.lucenetutorial.com/lucene-nrt-hello-world.html</a> )
Basic Concepts ( <a href="http://www.lucenetutorial.com/basic-concepts.html">http://www.lucenetutorial.com/basic-concepts.html</a> )
Lucene vs Solr ( <a href="http://www.lucenetutorial.com/lucene-vs-solr.html">http://www.lucenetutorial.com/lucene-vs-solr.html</a> )
Lucene Query Syntax ( <a href="http://www.lucenetutorial.com/lucene-query-syntax.html">http://www.lucenetutorial.com/lucene-query-syntax.html</a> )
Lucene Query Builder ( <a href="http://www.lucenetutorial.com/lucene-query-builder.html">http://www.lucenetutorial.com/lucene-query-builder.html</a> )
Your First Project ( <a href="http://www.lucenetutorial.com/your-first-project.html">http://www.lucenetutorial.com/your-first-project.html</a> )

## Techniques

Sample Apps
<ul style="list-style-type: none"> <li>- TextFileIndexer.java (<a href="http://www.lucenetutorial.com/sample-apps/textfileindexer-java.html">http://www.lucenetutorial.com/sample-apps/textfileindexer-java.html</a>)</li> </ul>
Techniques
<ul style="list-style-type: none"> <li>- Permission Filtering (<a href="http://www.lucenetutorial.com/techniques/permission-filtering.html">http://www.lucenetutorial.com/techniques/permission-filtering.html</a>)</li> <li>- Lucene+PHP (<a href="http://www.lucenetutorial.com/techniques/lucene-php.html">http://www.lucenetutorial.com/techniques/lucene-php.html</a>)</li> <li>- Indexing Databases (<a href="http://www.lucenetutorial.com/techniques/indexing-databases.html">http://www.lucenetutorial.com/techniques/indexing-databases.html</a>)</li> </ul>
Advanced Topics
<ul style="list-style-type: none"> <li>- Scoring (<a href="http://www.lucenetutorial.com/advanced-topics/scoring.html">http://www.lucenetutorial.com/advanced-topics/scoring.html</a>)</li> </ul>

## Misc

Need help with Lucene? ( <a href="http://www.lucenetutorial.com/consulting.html">http://www.lucenetutorial.com/consulting.html</a> )
About Me ( <a href="http://www.supermind.org/bio.html">http://www.supermind.org/bio.html</a> )

## Siblings

SolrTutorial.com ( <a href="http://www.solrtutorial.com">http://www.solrtutorial.com</a> )
ElasticSearchTutorial.com ( <a href="http://www.elasticsearchtutorial.com">http://www.elasticsearchtutorial.com</a> )
Solr vs ElasticSearch ( <a href="http://solr-vs-elasticsearch.com">http://solr-vs-elasticsearch.com</a> )

© Copyright 2017 Kelvin Tan - Lucene and Solr consultant (<http://www.supermind.org/>)  
Sponsored by Flexile Vertical Search Crawler (<http://verticalsearchcrawler.com/>)

# Programmatic construction of queries

Lucene queries can also be constructed programmatically. This can be really handy at times. Besides, there are some queries which are not possible to construct by parsing.

Available query objects as of 3.4.0 are:

- BooleanQuery
- ConstantScoreQuery
- CustomScoreQuery
- DisjunctionMaxQuery
- FilteredQuery
- MatchAllDocsQuery
- MultiPhraseQuery
- MultiTermQuery
- PhraseQuery
- RangeQuery
- SpanQuery
- TermQuery
- ValueSourceQuery

Use the BooleanQuery object to join and nest queries.

These classes are part of the org.apache.lucene.search package  
([http://lucene.apache.org/java/3\\_5\\_0/api/core/org/apache/lucene/search/package-summary.html](http://lucene.apache.org/java/3_5_0/api/core/org/apache/lucene/search/package-summary.html)).

Here's a simple example:

```
String str = "foo bar";
String id = "123456";
BooleanQuery bq = new BooleanQuery();
Query query = qp.parse(str);
bq.add(query, BooleanClause.Occur.MUST);
bq.add(new TermQuery(new Term("id", id), BooleanClause.Occur.MUST_NOT);
```

28 Comments

Lucene Tutorial

Login

Recommend 30

Share

Sort by Best



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS ?

Name



Spieraccioni • 6 years ago

in the last sample the last line of code is wrong due to the lack of a right closing parenthesis

33 ^ | v • Reply • Share

iroshan priyantha • 5 years ago

Hi I'm doing a web project & it has a function to find the similarity between two documents. these documents are text files. So i need to use cosine similarity or lucene to do this?i couldnt find any code sample for this. What i need is code sample for finding document similarity using lucene or cosine similarity.Please help me to overcome this problem. (i'm developing the program using java.please provide me complete code sample for this). thanks in advance.

5 ^ | v • Reply • Share



Funf → iroshan priyantha • 5 years ago

Spoonfeed me, please.

54 ^ | v • Reply • Share

Malcolm Greaves → iroshan priyantha • 4 years ago

Treat each word as a vector dimension. Each document is a vector; the values on each dimension are the word count. Cosine similarity is straightforward from this representation.

BTW this is called a bag of words model:

[http://en.wikipedia.org/wiki/Bag\\_of\\_words\\_model](http://en.wikipedia.org/wiki/Bag_of_words_model)

2 ^ | v • Reply • Share

**David\_Throop** • 5 years ago

May the indexed tokens include punctuation characters? I need to allow users to search for requirement numbers of the form  
V.SFR\_TM.262

I *think* what's happening is that Lucene is breaking the requirement number into 4 separate tokens. Can I control the tokenization process?

2 ^ | v • Reply • Share ›

**benafia salem** → David\_Throop • 5 years ago

Hello David. If your field is tokenized, let's say using StandardAnalyzer, the stored value would be for that sample : vsfrtm262... right?

Now if someone needs to find it, input can be vs.fr.Tm\_262 or whatever combination of punctuation. that's the interest of indexing your data.

If you want them to put the exact value to find that entry, you may try the whitespace analyzer (which uses a whitespace tokenizer). There are lots of analyzers you may test using Luke to choose what fits your need.

^ | v • Reply • Share ›



**Sss** • 6 years ago

ty

2 ^ | v • Reply • Share ›

**Brian Yuan** • 2 years ago

Hi Everyone,

Is there a maximum length limitation for the exact match? I tried to search a document that contains a sentence of "An investment in a money market fund is not insured or guaranteed by the Federal Deposit Insurance" but got the following result:

"An investment in a money market fund is not" - OK  
 "An investment in a money market fund is not insured" - No  
 "is not insured or guaranteed by the Federal Deposit" - OK  
 "fund is not insured or guaranteed by the Federal Deposit" - No  
 "is not insured or guaranteed by the Federal Deposit Insurance" - No

If there is a limitation to the length of the exact match, can I change it?

Thanks,

Brian

1 ^ | v • Reply • Share ›

**Tahir Shahzad** • 4 years ago

i urgently need to use lucence with my php project. can anyone guide me step by step? how can i configure it and perform query?

1 ^ | v • Reply • Share ›



**kjhgfjfx** → Tahir Shahzad • 3 years ago

No.

10 ^ | v • Reply • Share ›



**Jamal** → Tahir Shahzad • 3 years ago

I have done a implementation with Lucene for web site search. Check [jamalmohideen.com/isearch/s...](http://jamalmohideen.com/isearch/s...), contact me if you still need help.

^ | v • Reply • Share ›

**mehwish** → Jamal • 2 years ago

hi i need your help .can you please provide me sample code to search through lucene

^ | v • Reply • Share ›

**Shabir Ahmad Bhat** • 3 months ago

how can i give this query (title:"foo bar" AND body:"quick fox") OR title:fox directly

i have queryString = "(title:"foo bar" AND body:"quick fox") OR title:fox "

how i make Query object of this String

can any one help me. i tried this but i not works:

QuervParser parser = new QuervParser(null analyzer);

```
Query queryParser = new QueryParser(new Analyzer(),
Query query = parser.parse(queryString);
TopDocs hits = searcher.search(query, 1);
```

^ | v • Reply • Share ›

**杨煜滨** → Shabir Ahmad Bhat • 3 months ago

I have the same question

^ | v • Reply • Share ›

**kostali hassan** • a year ago

I try to index rich data for exemple CVs msword and pdf ; how i can build a query to get e-mail of such CV who is in the content of document. build query to serach e-mail in the field content of such file.

^ | v • Reply • Share ›

**PS Hayer** • 2 years ago

is there any way to sort results based on a date type field. I want to sort results Descending order?

^ | v • Reply • Share ›

**Maynou-Taec** → PS Hayer • 2 years ago

hello i'm sorry but me too i'm trying to do the same thing and i just began to deal with lucene so can i ask for help if u figure it out. I have a question: do we need to add the field date & title for exp in the index & in the research procudure we will have to sort the result by the creation date for exp. does this work? if u have any intresting link would u plz give them to me. Thank u

^ | v • Reply • Share ›

**PS Hayer** → Maynou-Taec • 2 years ago

Hi, It is been a while now so I vaguely remember that there was no way to sort results based on 'Created On' Date. I was working with ADX Studio and it was really a pain sort results. So finally did it the dirty way. Picked up all the results in Lucene query (in our case there were around 4000 results in total), then run a sort in your code and handle the Paging yourself. I don't exactly remember but I am sure that there was no way to sort results using Lucene. But you can pick the results in between two certain dates.

^ | v • Reply • Share ›



**Evandro Oliveira Calazans** • 2 years ago

What's version of lucene-analyzers-common-X.X.X.jar, lucene-core-X.X.X.jar, lucene-queryparser-X.X.X.jar and pdfbox-app-X.X.X.jar which can i use with JDK/JRE 1.5?

^ | v • Reply • Share ›

**Urvashi An Extrovert** • 3 years ago

how can i find a string using term-query .....in child bean..i able to find it in parentbean,but not in childbean.../

^ | v • Reply • Share ›

**ClintJCL** • 4 years ago

I have queries that look like (5W(something\*, ( (blah\* W bla OR (orange\* W red\*) OR (with W other\* W in W error\* W play W to W atari)

The co-workers here say this is Lucene, but it looks like nothing listed here.

So what is it?

^ | v • Reply • Share ›



**James B. Helix Roveda** • 4 years ago

It would really help to actually verbalize what the simple example is doing. Because I am not sure. So, I guess i ain't that simple after all.

^ | v • Reply • Share ›

**bob** • 4 years ago

we are trying to use lucene to differentiate between postcodes but are having problems once the postcode goes past 10. for example cm1\* and cm10\* can't co-exist is a search query, or can they? all help appciated!

^ | v • Reply • Share ›

**Dheeraj** • 5 years ago

Hi.. I m new to the lucene.. I m doing an E-Commerce application.. Given I have a main category called "Organic

Produce" and a Sub Category called "Organic fruits" under another main category called "Fruits".. Both "Organic Produce" and "Organic fruits" contains some list of products.. If in search query if i entered "Organic fruits".. I need to display products that belong to "Organic fruits".. But here i m getting products from both "Organic Produce" and "Organic fruits".. I m using MultiFieldQueryParser.. Is there any alternate queryparser to search.. If so.. how.. Can anybody Please suggest any solution for this..

^ | v • Reply • Share ›

**sash** • 5 years ago

hi, how can find all documents having two different terms in them anywhere. e.g. Say all documents having "tree" and "computer" in their contents. I have constructed an inverted index of all document contents. It is like intersection of docs having "tree" AND "Computer" .

^ | v • Reply • Share ›

**Nasrulla Kandukuri** • 5 years ago

Hi Team,

I am new to Lucene. I have a requirement like, I want to get the products like whose product name starts with A or a. My Product names are as follows

- 1) Green Apple
- 2) Blue Apple
- 3) Apple
- 4) Orange Apple

And I framed a query like 'title:A\*', According to my assumption it has to display only one i.e Apple. But it is giving all 4 products. How could I achieve this requirement. Please suggest me something.

^ | v • Reply • Share ›



**floyd** → **Nasrulla Kandukuri** • 5 years ago

u should take the whole phrase as a term, which means u must not analyze the field as following:

```
Field f1 = new Field("content", "Blue Apple", Field.Store.YES, Field.Index.NOT_ANALYZED);  
try
```

2 ^ | v • Reply • Share ›

**Shabir Ahmad Bhat** → **floyd** • 3 months ago

how can i give this query (title:"foo bar" AND body:"quick fox") OR title:fox directly  
i have queryString = "(title:"foo bar" AND body:"quick fox") OR title:fox "  
how i make Query object of this String

^ | v • Reply • Share ›