

Ten. Million. Questions. Let's celebrate [all we've done together](#).

Stack Overflow is a question and answer site for professional and enthusiast programmers. It's 100% free.

[Take the 2-minute tour](#) ×

Full text search on Neo4j over rich text with html markup



Work at Smartsheet

Bellevue, WA



Alex Vorobiev



Slaven Slugic

▲ In my Neo4j application I have a `Product` entity with a `name` and `description` fields. Both of these fields are used in legacy indexing over Lucene.

▼ `Product.name` is a simple text and there are no issues here but `Product.description` can contain HTML markup and elements.

★ Right now for my index I use `StandardAnalyzer(Version.LUCENE_36)`. What analyzer should I use in order to skip all HTML elements ?

1

How to tell Neo4J Lucene index to not use any HTML elements in `Product.description` ? I'd like to index only words.

UPDATED:

I have found following class `HTMLStripCharFilter` and reimplemented my Analyzer as following:

```
public final class StandardAnalyzerV36 extends Analyzer {  
    private Analyzer analyzer;  
  
    public StandardAnalyzerV36() {  
        analyzer = new StandardAnalyzer(Version.LUCENE_36);  
    }  
  
    public StandardAnalyzerV36(Set<?> stopWords) {  
        analyzer = new StandardAnalyzer(Version.LUCENE_36, stopWords);  
    }  
  
    @Override  
    public final TokenStream tokenStream(String fieldName, Reader reader) {  
        return analyzer.tokenStream(fieldName, new HTMLStripCharFilter(CharReader.ge  
    }  
  
    @Override  
    public final TokenStream reusableTokenStream(String fieldName, Reader reader) th  
        return analyzer.reusableTokenStream(fieldName, reader);  
    }  
}
```

also I have added a new maven dependency to my Neo4j project:


```
<dependency>  
    <groupId>org.apache.lucene</groupId>  
    <artifactId>lucene-analyzers-common</artifactId>  
    <version>3.6.2</version>  
</dependency>
```

asked 4 months ago

viewed 44 times

active 17 days ago

Blog

 [Why Stack Overflow is a Good Workplace for Women](#)



Smartsheet

based in Bellevue, WA

Who you'll work with...



Alex Vorobiev



Slaven Slugic



Kyan Skeem

[view company page](#)

Related

- 0 Wildcard search with cypher query and spring neo4j
- 0 Spring Data Neo4j search and index
- 3 Full text search in Neo4j with spaces
- 1 neo4j not starting after disk full
- 1 Neo4j index for full text search
- 1 Neo4j 2.1.5 incremental backup fails but full backup succeeds
- 0 Full Text indexing in Neo4j with Spring Data Neo4j 3.2.0

```
<artifactId>lucene-analyzers</artifactId>
<version>3.6.2</version>
</dependency>
```

Everything works fine right now, but I'm not sure that the method

```
@Override
public final TokenStream tokenStream(String fieldName, Reader reader) {
    return analyzer.tokenStream(fieldName, new HTMLStripCharFilter(CharReader.ge
}
```

is a proper place for `HTMLStripCharFilter` initialization.


Please correct me if I'm wrong.

lucene neo4j spring-data-neo4j

share improve this question

edited Apr 4 at 15:24

asked Apr 4 at 14:47

 alexanoid
1,008 ● 12 ● 28

add a comment

1 Answer

active oldest votes

I have added following init method:

```
@PostConstruct
public void init() {
    GraphDatabaseService graphDb = template.getGraphDatabaseService();
    try (Transaction t = graphDb.beginTx()) {
        Index<Node> autoIndex = graphDb.index().forNodes("node_auto_index");
        graphDb.index().setConfiguration(autoIndex, "type", "fulltext");
        graphDb.index().setConfiguration(autoIndex, "to_lower_case", "true");
        graphDb.index().setConfiguration(autoIndex, "analyzer", StandardAnalyzer
t.success();
    }
}
```

and created following class:

```
public final class StandardAnalyzerV36 extends Analyzer {

    private Analyzer analyzer;

    public StandardAnalyzerV36() {
        analyzer = new StandardAnalyzer(Version.LUCENE_36);
    }

    public StandardAnalyzerV36(Set<?> stopWords) {
        analyzer = new StandardAnalyzer(Version.LUCENE_36, stopWords);
    }


    @Override
    public final TokenStream tokenStream(String fieldName, Reader reader) {
        return analyzer.tokenStream(fieldName, new HTMLStripCharFilter(CharReader.ge
    }

    @Override
    public final TokenStream reusableTokenStream(String fieldName, Reader reader) th
        return analyzer.reusableTokenStream(fieldName, reader);
    }
}
```

Now everything works as expected. Hope it will help someone else. Good luck.

share improve this answer

answered Aug 2 at 16:41

 alexanoid
1,008 ● 12 ● 28

add a comment


0 Which type of indexing is required in Neo4j for search from many type of Labels in Spring Data Neo4j

0 How to write dynamic search query for Spring-Data-Neo4j

0 Neo4j / Spring data Full-text search all fields functionality?

Hot Network Questions

 Religion After the Discovery of a Multiverse

 Prove the theorem on analytic geometry in the picture.

0 Does touching the tent outer wall from the inside make it leak?

ELL Is Let us = Let's?

Cy Combining several symmetric ciphers using XOR

PCG Construct a polynomial with given roots

WW Which type of drill bit for which jobs?


PCG The Programming Language Quiz

Preventing a Print expression from generating a new cell each time it executed

How to deal with people looking at their mobile phone during my presentation?

What is aliasing and what causes it?

Carrying a handgun in other countries with a US concealed carry permit

 Manager sounds upset every time I inform him of a (minor) obstacle

H Did a Kamikaze ever impact the hull of a ship, as opposed to the deck?

Did Arnold Rimmer kill the Red Dwarf crew?


Is it 'Batmen' or 'Batmans'?


 How does java.util.EnumSet<E> work?

Remove a language from the "Preferred language order"

How can I grant a non-Super User group permission to force check-in of articles?

Why don't decoupling/bypass capacitors need resistors to perform their function, like regular filters?

 Using siunitx, \ohm results in an italic Omega

 How do I know when my sauce is reduced enough?




MO Important formulas in Combinatorics

QF How do I calculate approximate equity liquidity?

Your Answer

Sign up or [log in](#)

Post as a guest

-  Sign up using Google
-  Sign up using Facebook
-  Sign up using Stack Exchange

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the [privacy policy](#) and [terms of service](#).

Not the answer you're looking for? Browse other questions tagged `lucene` `neo4j` `spring-data-neo4j` or [ask your own question](#).



TECHNOLOGY			LIFE / ARTS	CULTURE / RECREATION	SCIENCE	OTHER
Stack Overflow	Programmers	Database Administrators	Photography	English Language & Usage	Mathematics	Stack Apps
Server Fault	Unix & Linux	Drupal Answers	Science Fiction & Fantasy	Skeptics	Cross Validated (stats)	Meta Stack Exchange
Super User	Ask Different (Apple)		Graphic Design		Mi Yodeya (Judaism)	Theoretical Computer Science
Web Applications	WordPress Development	User Experience	Movies & TV	Travel	Physics	Stack Overflow Careers
Ask Ubuntu	Geographic Information Systems	Mathematica	Seasoned Advice (cooking)	Christianity	MathOverflow	
Webmasters	Electrical Engineering	Salesforce	Home Improvement	Arqade (gaming)	Chemistry	
Game Development		ExpressionEngine®		Bicycles	Biology	
TeX - LaTeX		Android Enthusiasts	Answers	Personal Finance & Money	Role-playing Games	more (5)
	Information Security	more (13)	Academia	more (21)		
			more (9)			