Questions

Tags

Users

Bellevue, WA

Badges

l Inanswered

Ask Ouestion

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**

Alex Vorobiev



Slaven Slugic

asked 4 months ago

viewed 44 times

active 17 days ago



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.

3

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 ?

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:

Blog

Why Stack Overflow is a Good Workplace for Women



Smartsheet

Who you'll work with... Alex Vorobiev



Slaven Slugic



Kyan Skeem



Related

- Wildcard search with cypher query and spring neo4i
- O Spring Data Neo4j search and index
- Full text search in Neo4j with spaces
- neo4j not starting after disk
- Neo4j index for full text search
- Neo4j 2.1.5 incremental backup fails but full backup succeeds
- Full Text indexing in Neo4j with Spring Data Neo4j 3.2.0



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

```
<dependency>
  <groupId>org.apache.lucene</groupId>
```

```
<artifactig>lucene-analyzers</artifactig>
<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



oldest

votes

active

add a comment

1 Answer

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) the 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

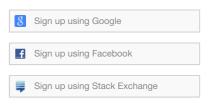
- Which type of indexing is required in Neo4j for search from many type of Labels in Spring Data Neo4j
- How to write dynamic search query for Spring-Data-Neo4i
- Neo4j / Spring data Fulltext search all fields functionality?

Hot Network Questions

- Religion After the Discovery of a Multiverse
- Prove the theorem on analytic geometry in the picture.
- Does touching the tent outer wall from the inside make it leak?
- Is Let us = Let's?
- Combining several symmetric ciphers using XOR
- Construct a polynomial with given roots
- Which type of drill bit for which jobs?
- 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
- Did a Kamikaze ever impact the hull of a ship, as opposed to the deck?
- Did Arnold Rimmer kill the Red
 Dwarf crew?
- (A) 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
- How do I calculate approximate equity liquidity?



Sign up or log in



Post as a guest

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.

a question feed

