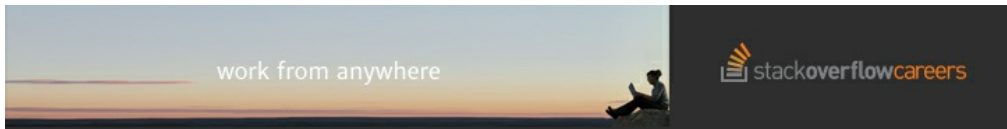


Ten. Million. Questions. Let's celebrate [all we've done together](#).

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

Take the 2-minute tour ×

## Neo4j: Special characters and time lag



**Goal:** I am trying to make a Neo4j instance of the DBLP database on the basis of the publicly available DBLP XML file available [here](#). I have modeled the database as a bipartite graph where the authors are in one set and the publications in the other set. To obtain all coauthors of John Doe one has to make the following Cypher query:

```
MATCH (a:Author)-[:WROTE]->(publication)-[:WROTE]-(b:Author) WHERE a.name = "John Doe"
RETURN DISTINCT b"
```

**Problem 1:** There seems to be a problem partly related to special characters, such as *ë*, *æ*, *í*, etc. When I, in my browser at the address <http://localhost:7474/browser/>, enter the query "MATCH (a:Author)-[:WROTE]->(p)-[:WROTE]-(b:Author) WHERE a.name = "Jan Arne Telle" RETURN DISTINCT b", I should get 58 unique results (coauthors), but I get 79 results. For instance, coauthor Daniël Paulusma is split into three results: "Dani", "ë", "I Paulusma". But in fact, I also get coauthor David Keldsen as three results: "David Keldsen", "David", and "Keldsen". So the problem is not only related to special characters.

**Problem 2:** Results for the above mentioned query were returned in 90697 ms.

EDIT: After making several such queries results are returned in 2000 ms to 4000 ms.

Here is all the code:

Entry point: Application.java:

```
package std;

import java.io.File;

import org.neo4j.graphdb.GraphDatabaseService;
import org.neo4j.graphdb.Transaction;
import org.neo4j.graphdb.factory.GraphDatabaseFactory;
import org.neo4j.kernel.impl.util.FileUtils;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.boot.CommandLineRunner;
import org.springframework.boot.SpringApplication;
import org.springframework.boot.autoconfigure.SpringBootApplication;
import org.springframework.context.annotation.Bean;
import org.springframework.context.annotation.Configuration;
import org.springframework.data.neo4j.config.EnableNeo4jRepositories;
import org.springframework.data.neo4j.config.Neo4jConfiguration;
import org.springframework.data.neo4j.core.GraphDatabase;

import java.io.FileInputStream;
import java.io.IOException;
import java.io.InputStream;

import javax.xml.parsers.ParserConfigurationException;
import javax.xml.parsers.SAXParser;
import javax.xml.parsers.SAXParserFactory;

import org.xml.sax.SAXException;
import org.apache.xerces.util.SecurityManager;

@SpringBootApplication
```

asked 4 months ago

viewed 47 times

active 4 months ago

### Blog

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








### Linked

0 [SAX parsing and special characters](#)

### Related

- 6 [Issue wiring together spring-data-neo4j standalone spring project](#)
- 3 [Spring/Neo4J - Is impermanentGraphDatabase compatible with Transaction Annotation?](#)
- 0 [Spring Neo4j Data is destroying db on every deploy](#)
- 0 [zookeeper ConnectException in embedded neo4j](#)
- 0 [Spring Neo4j batch over REST](#)
- 0 [After neo4j backup no index hit](#)
- 3 [I can't update spring data neo4j version](#)
- 1 [Spring Data Neo4J @Query](#)
- 0 [Spring neo4j cannot add relationships](#)
- 0 [Spring data neo4j](#)

### Hot Network Questions

-  [I added more water to my ciabatta dough by accident](#)
-  [Can a plane fly backward if its prop's in reverse?](#)
-  [Did Arnold Rimmer kill the Red Dwarf crew?](#)
-  [Could a communication satellite that fails to reach GEO be repurposed?](#)
-  [Approximation of Borel sets by a countable collection of majorants](#)
-  [Password strength of random words and brainflayer](#)
-  [Legacy application unreachable code](#)

```

public class Application implements CommandLineRunner {

    @Configuration
    @EnableNeo4jRepositories(basePackages = "std")
    static class ApplicationConfig extends Neo4jConfiguration {

        public ApplicationConfig() {

```

Author.java:

```

package std;

import org.springframework.data.neo4j.annotation.GraphId;
import org.springframework.data.neo4j.annotation.Indexed;
import org.springframework.data.neo4j.annotation.NodeEntity;
import org.springframework.data.neo4j.annotation.Query;
import org.springframework.data.neo4j.support.index.IndexType;

@NodeEntity
public class Author {

    @GraphId
    private Long id;

    @Indexed(indexName = "names", unique = true, indexType = IndexType.FULLTEXT)
    private String name;

    public Author() {
    }

    public Author(String name) {
    }

    @Override
    public boolean equals(Object obj) {
        if (this == obj)
            return true;


















        if (obj == null)
            return false;

        if (this.getClass() != obj.getClass())
            return false;

        Author other = (Author) obj;

```

Publication.java:

-  Construct a polynomial with given roots
-  Can a Glaive be stowed in a Ruby Scabbard?
-  Why are my tomatoes cracking and what can I do?
-  Preventing a Print expression from generating a new cell each time it executed
-  How do I calculate approximate equity liquidity?
-  Remove a language from the "Preferred language order"
-  Should you always minimize cognitive load
-  Was フリスマスケーキ used metaphorically about women?
-  Seven overlapping circles
-  Does developing on ARM Processors from different vendors differ much?
-  Two definitions of Green's function
-  When does the designer pay for a reprint
-  Weird performance increase in simple benchmark
-  As a professional photographer, how can one handle a wedding photographer being rude or demanding in a way that will compromise overall image quality?
-  Dullness vs. going overboard: Should I be calling people 'enfants terribles' in an academic paper?
-  Aerofoil that gives reasonably good lift for both flow directions: forward and backward
-  The Programming Language Quiz

```

package std;

import java.io.Serializable;
import java.util.HashSet;
import java.util.Set;

import org.neo4j.graphdb.Direction;
import org.springframework.data.neo4j.annotation.GraphId;
import org.springframework.data.neo4j.annotation.Indexed;
import org.springframework.data.neo4j.annotation.NodeEntity;
import org.springframework.data.neo4j.annotation.RelatedTo;
import org.springframework.data.neo4j.support.index.IndexType;

@NodeEntity
public class Publication implements Serializable {

    private static final long serialVersionUID = -6393545300391560520L;

    @GraphId
    Long nodeId;

    private String type = "";
    private String key = "";
    private String mdate = "";
    private String pubtype = "";
    private String reviewid = "";
    private String rating = "";

    @RelatedTo(type = "WROTE", direction = Direction.INCOMING)
    private Set<Author> authors = new HashSet<Author>();
    private String editor = "";

    @Indexed(indexType = IndexType.FULLTEXT, indexName = "titles")
    private String title = "";

    private String booktitle = "";
    private String pages = "";
    private String year = "";

```

PublicationRepository.java:

```

package std;

import org.springframework.data.neo4j.repository.GraphRepository;

public interface PublicationRepository extends GraphRepository<Publication> {

    Publication findByTitle(String title);

}

```

SaxHandler.java:

```

package std;

import java.util.ArrayList;
import java.util.List;
import java.util.Stack;

import org.neo4j.graphdb.Transaction;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.data.neo4j.core.GraphDatabase;
import org.xml.sax.Attributes;
import org.xml.sax.SAXException;
import org.xml.sax.helpers.DefaultHandler;

public class SaxHandler extends DefaultHandler {

    private Stack<String> qNameStack = new Stack<String>();
    private Stack<Publication> publicationStack = new Stack<Publication>();
    private String publicationType = null;
    private PublicationRepository publicationRepository = null;
    private Publication publication = null;
    private Author author = null;
    private String currentElement = null;
    private String value = null;
    private boolean insideTitle = false;

    private GraphDatabase graphDatabase;

```

```

private Transaction tx = null;

private static int counter = 0;

public List<Publication> getPublications() {
    return publications;
}

@Autowired
public SaxHandler(PublicationRepository publicationRepository, GraphDatabase g
    this.publicationRepository = publicationRepository;

```

pom.xml:

```

<?xml version="1.0" encoding="UTF-8"?>
<project xmlns="http://maven.apache.org/POM/4.0.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apa

    <modelVersion>4.0.0</modelVersion>

    <groupId>com.dblp</groupId>
    <artifactId>graphdbcreator</artifactId>
    <version>0.1.0</version>

    <parent>
        <groupId>org.springframework.boot</groupId>
        <artifactId>spring-boot-starter-parent</artifactId>
        <version>1.2.2.RELEASE</version>
    </parent>

    <dependencies>
        <dependency>
            <groupId>org.springframework.boot</groupId>
            <artifactId>spring-boot-starter</artifactId>
        </dependency>
        <dependency>
            <groupId>org.springframework</groupId>
            <artifactId>spring-context</artifactId>
        </dependency>
        <dependency>
            <groupId>org.springframework</groupId>
            <artifactId>spring-tx</artifactId>
        </dependency>
        <dependency>
            <groupId>org.springframework.data</groupId>
            <artifactId>spring-data-neo4j</artifactId>
        </dependency>
        <dependency>
            <groupId>org.hibernate</groupId>
            <artifactId>hibernate-validator</artifactId>

```

spring neo4j sax saxparser spring-data-neo4j

share improve this question

edited Apr 4 at 5:45

asked Apr 4 at 5:12



add a comment

2 Answers

active oldest votes



0

For problem 1 try to setup a manual index with an analyzer fitting your needs. See <http://blog.armbruster-it.de/2014/10/deep-dive-on-fulltext-indexing-with-neo4j/> for details on how to use custom analyzers.



Another option would be to use stemming logic on application side and store the stemmed name in a secondary property.

A third option is adding "SIMILAR" relationships between author nodes referring to the very same person.


Regarding problem 2: make sure to have an index on the name property for authors:

```
CREATE INDEX ON :Author(name)
```

The difference on query times for subsequent calls is easily explained with caching, read more on <http://neo4j.com/docs/stable/configuration-caches.html>

share improve this answer

answered Apr 4 at 7:57

 **Stefan Armbruster**  
23.1k ● 2 ● 28 ● 51

I will look further at your solutions to problem 1. Concerning your solution to problem 2, do I need this when I provide the annotation @Indexed(indexName = "names", unique = true, indexType = IndexType.FULLTEXT) on top of the name property in the Author class? – [GCOverhead](#) Apr 4 at 10:35

for problem2: just use @Indexed(unique=true) on property name. – [Stefan Armbruster](#) Apr 4 at 10:42

add a comment



0



It seems my SAX handler was flawed. For instance, given a tag `<author>Daniël Paulusma</author>`, the parser would make one call to the `characters()` method for "Dani", another call to `characters()` for "ël", and a third call to `characters()` for "l Paulusma". I found a simple solution to this problem here: [SAX parsing and special characters](#).

share improve this answer

answered Apr 6 at 5:33


 **GCOverhead**  
3 ● 2


add a comment


Your Answer

**B** *I*          

Sign up or [log in](#)

 Sign up using Google

 Sign up using Facebook

 Sign up using Stack Exchange

Post as a guest

**Name**

**Email**

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 [spring](#) [neo4j](#) [sax](#) [saxparser](#) [spring-data-neo4j](#) or [ask your own question](#).



[tour](#) [help](#) [blog](#) [chat](#) [data](#) [legal](#) [privacy policy](#) [work here](#) [advertising info](#) [mobile](#) [contact us](#) [feedback](#)

#### TECHNOLOGY

Stack Overflow  
Server Fault  
Super User  
Web Applications  
Ask Ubuntu  
Webmasters  
Game Development  
TeX - LaTeX

Programmers  
Unix & Linux  
Ask Different (Apple)  
WordPress Development  
Geographic Information Systems  
Electrical Engineering  
Android Enthusiasts  
Information Security

Database Administrators  
Drupal Answers  
SharePoint  
User Experience  
Mathematica  
Salesforce  
ExpressionEngine® Answers  
**more (13)**

#### LIFE / ARTS

Photography  
Science Fiction & Fantasy  
Graphic Design  
Movies & TV  
Seasoned Advice (cooking)  
Home Improvement  
Personal Finance & Money  
Academia  
**more (9)**

#### CULTURE / RECREATION

English Language & Usage  
Skeptics  
Mi Yodeya (Judaism)  
Travel  
Christianity  
Arqade (gaming)  
Bicycles  
Role-playing Games  
**more (21)**

#### SCIENCE

Mathematics  
Cross Validated (stats)  
Theoretical Computer Science  
Physics  
MathOverflow  
Chemistry  
Biology  
**more (5)**

#### OTHER

Stack Apps  
Meta Stack Exchange  
Area 51  
Stack Overflow Careers