

To be or not to be.

Neo4j Full Text Search Tips and Tricks

Christophe Willemsen
CTO - GraphAware
@ikwattro

#NODES19 - Oct 10, 2019

 graphaware.com

 [@graph_aware](https://twitter.com/graph_aware)



GraphAware®

Agenda

- Full Text Search Syntax refresher
- Relevant Search using Graphs
- Advanced Full Text Search Queries



How is text data indexed?

0. Mastering Highly Distributed Architecture
 1. Graph Algorithms in Neo4j : All Pairs Shortest Path
 2. Announcing the Neo4j GraphTour in 8+ Cities Worldwide
 3. Graph Algorithms in Neo4j : Single Source Shortest Path

Term dictionary

mastering	0
highly	1
distributed	2
architecture	3
graph	4
algorithms	5
in	6
neo4j	7
all	8
pairs	9
shortest	10
path	11
announcing	12
the	13
graphtour	14
8+	15
cities	16
worldwide	17
single	18
source	19

Postings List

0	[0]
1	[0]
2	[0]
3	[0]
4	[1,3]
5	[1,3]
6	[1,2,3]
7	[1,2,3]
8	[1]
...	



<h1>The GraphAware's fifth year anniversary at the Prague office in Žitná</h1>

HTML character filter

The GraphAware's fifth year anniversary at the Prague office in Žitná

ASCII character collapsing filter

The GraphAware's fifth year anniversary at the Prague office in Zitna

Standard Tokenizer

The GraphAware's fifth year anniversary at the Prague office in Zitna

Lowercase token filter

the graphaware's fifth year anniversary at the prague office in zitna

Stopword token filter

graphaware's fifth year anniversary prague office zitna

Possessive token filter

graphaware fifth year anniversary prague office zitna



Default Stopwords for English

"a", "an", "and", "are", "as", "at", "be", "but",
"by", "for", "if", "in", "into", "is", "it",
"no", "not", "of", "on", "or", "such",
"that", "the", "their", "then",
"there", "these",
"they", "this", "to", "was", "will", "with"



Default Stopwords for English

"a", "an", "and", "are", "as", "at", **"be"**,
"but", "by", "for", "if", "in", "into", "is", "it",
"no", **"not"**, "of", "on", **"or"**, "such",
"that", "the", "their", "then",
"there", "these",
"they", "this", **"to"**, "was", "will", "with"



Custom Lucene Analyzer

```
package com.graphaware.neo4j;

import org.apache.lucene.analysis.Analyzer;
import org.apache.lucene.analysis.standard.StandardAnalyzer;
import org.apache.lucene.analysis.util.CharArraySet;
import org.neo4j.graphdb.index.fulltext.AnalyzerProvider;
import org.neo4j.helpers.Service;

@Service.Implementation(AnalyzerProvider.class)
public class StandardNoStopwordsAnalyzer extends AnalyzerProvider {

    public static final String STANDARD_ANALYZER_NAME = "standard-no-stopwords";

    public StandardNoStopwordsAnalyzer() {
        super(STANDARD_ANALYZER_NAME);
    }

    @Override
    public Analyzer createAnalyzer() {
        return new StandardAnalyzer(CharArraySet.EMPTY_SET);
    }

    @Override
    public String description() {
        return "The default, standard analyzer but without stopwords. Tokenizes on non-letter and filters out English stop words and punctuation. " +
            "Does no stemming, but takes care to keep likely product names, URLs and email addresses as single terms.";
    }
}
```



Custom Lucene Analyzer

```
ikwattro@mbp666 ~/d/_/nodes19> ll plugins/  
total 28264  
drwxr-xr-x  4 ikwattro  staff  128B Oct 10 09:46 .  
drwxr-xr-x 17 ikwattro  staff  544B Sep 22 20:25 ..  
-rw-r--r--@ 1 ikwattro  staff  13M Sep 22 20:30 apoc-3.5.0.4-all.jar  
-rw-r--r--  1 ikwattro  staff  4.9K Sep 23 12:08 fts-extra-1.0.0.jar
```



Relevant Search



Relevant Search

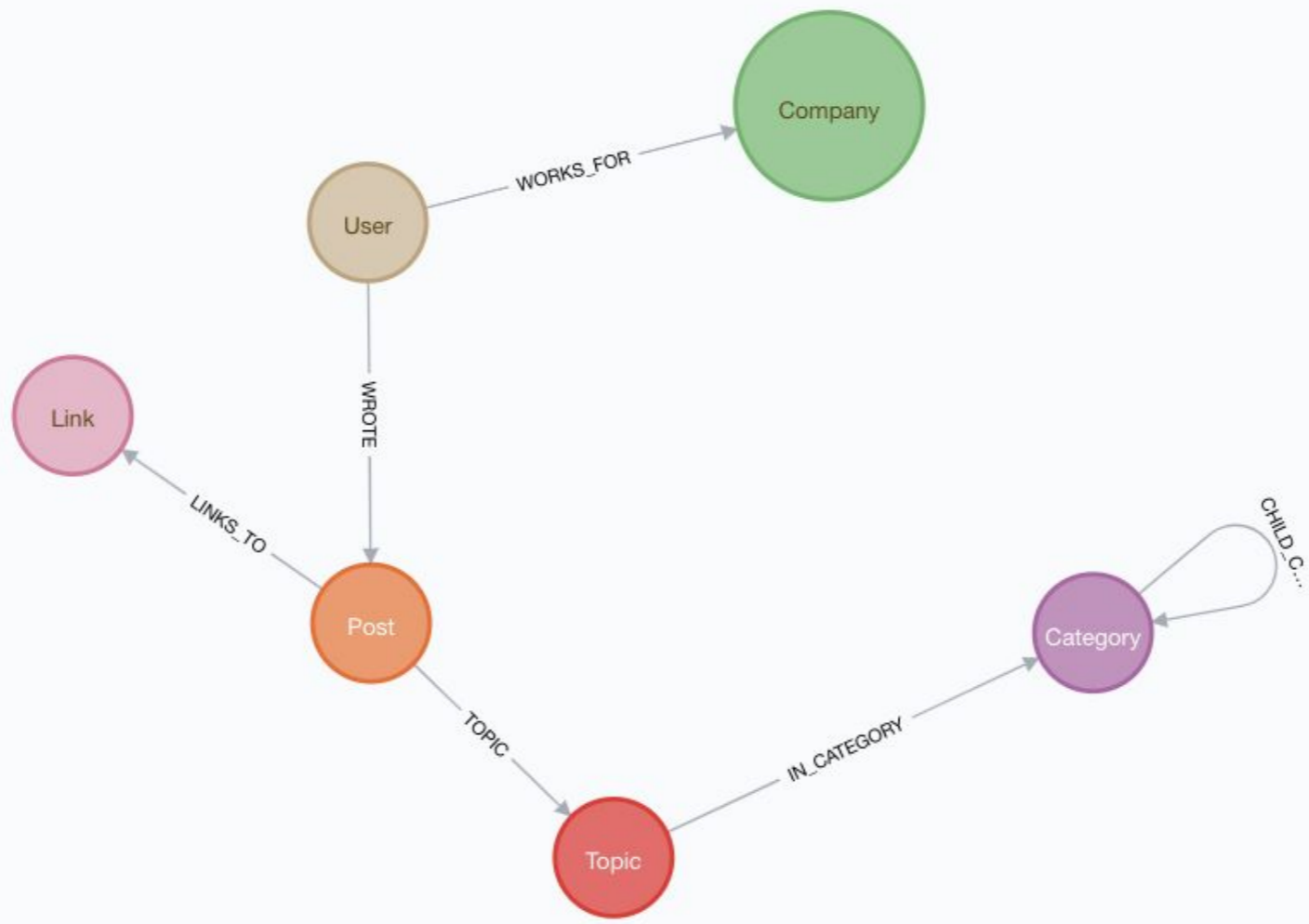
- You first query the graph to build the user context
- You then use the context to make search results relevant not only based on their text similarity, but also **based on who the user is in the network.**



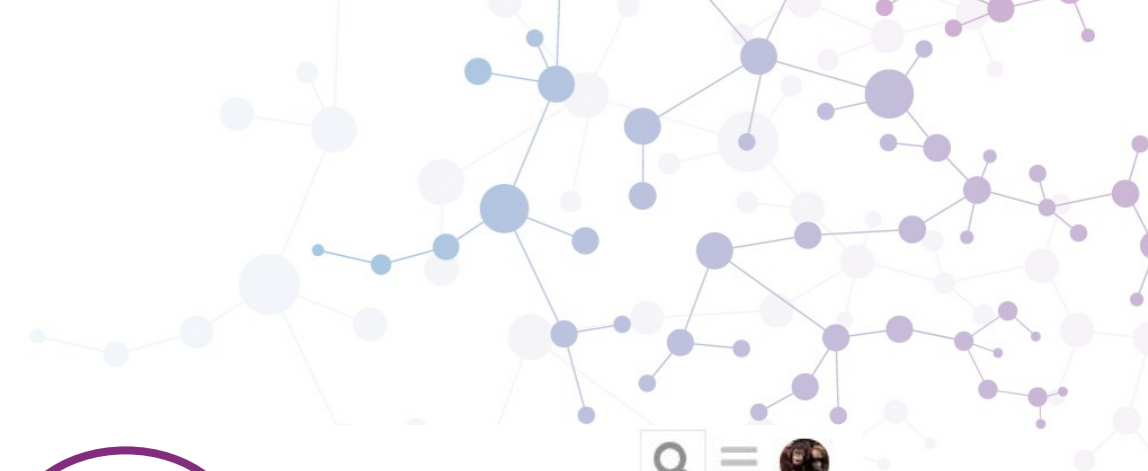
Relevant Search



- * (7)
 - Company (1)
 - Category (1)
 - User (1)
 - Post (1)
 - Talk (1)
 - Topic (1)
 - Link (1)
-
- * (6)
 - CHILD_CATEGORY (1)
 - LINKS_TO (1)
 - WROTE (1)
 - WORKS_FOR (1)
 - TOPIC (1)
 - IN_CATEGORY (1)



Relevant Search



Welcome to the Neo4j Community

Discussions Help by answering Get Certified Neo4j Speaker Program Graph Gear

Featured Community Member

Dr Estelle Scifo

Lead Data Scientist R&D at Motion-S



Featured Community Member
5th October 2019

Popular Community Content

graph modeling, when to use a property, a label, and a separate node?

g A Photo-Based Personalized Recommendations Application

Popular Community Projects

Halin v0.11 is now available! (The Trivium release)

Football exploration with Neo4r

Jan

options

- Jan**
jan
- Janakiramram...**
Ram
- L A Janota**
l.a.janota
- Pankaj Jangid**
pankaj.jangid
- James Janeway**
james.janeway

More...

Janos Szendi-Varga - Senior Consultant at GraphAware
Introduce-Yo... graphaware
Aug '18 - I worked in an innovation lab in Hungary a few years ago when I first met with Neo4j, and I started to learn this. It started with a few pet projects and meetups, I did the neo4j certification amon...

Janos - Discovering my Community Connections!
Random contest
Dec '18 - I'm Janos and here are my findings on the Neo4j Community Site: It is very cool, that there is a https://co...

Permission denied when restarting core servers in a casual cluster
Docker
May 3 - ...neo4j neo4j 4096 May 3 19:14 neo4j-docker work/neo4j-docker: total 28 drwxrwxrwx. 5 neo4j neo4j 39 Jan 17 21:17 core1 drwxrwxrwx. 5 neo4j neo4j 39 Jan 17 21:18 core2 drwxrwxrwx. 5 neo4j neo4j 39 Jan 17...

O'Reilly Open Source Convention (OSCON) - July 15/18 Portland, Oregon (CFP deadline January 15, 2019!)
Speaking Op... us,conference
Nov '18 - O'Reilly Open Source Convention (OSCON) - July 15/18 Portland, Oregon Open Source: Open source is how software is developed today – beginning with the flow of open ideas that ultimately leads to wo...

Family tree relationship mapping
Cypher
Jan 10 - ...ther someone has already mapped family tree relationships? E.g. create (Joe:Male { name: 'Joe' }), (Jane:Female { name: 'Jane' }), (Joe)-[:PARENT]-> (Jane), (Dave:Male { name: 'Dave' }), (Jane)-[:Parent...

More...



GraphAware®

Relevant Search




Welcome to the Neo4j community

Discussions Help by answering Get Certified Neo4j Speaker Program Graph Gear

Featured Community Member

Dr Estelle Scifo
Lead Data Scientist R&D at Motion-S



Featured Community Member
5th October 2019

Search: full text

- Putting full-text indexes on properties
- Full-text search on all nodes and all attributes
- Deep Dive into Neo4j 3.5 Full Text Search**
- How to use a full text index with different lucene analyzers for indexing and searching in neo4j?
- Question about passing parameter list into full text index

Popular Community Content

When graph modeling, when to use a property, a label, and a separate node?

Building A Photo-Based Personalized Recommendations Application

Popular Community Projects

Halin v0.11 is now available! (The Trivium release)

Football exporation with Neo4r

New Certified

Raavva

Caojingwei



GraphAware®

Hunger Games Questions for "To be or not to be. Full Text Search tips and tricks"

1. **Easy:** What type of properties can we index for full text search.
 - a. All properties
 - b. Only strings and integers
 - c. Only strings
2. **Medium:** How can I index stopwords such that they can be searched?
 - a. Using an analyzer available in Neo4j
 - b. Using a custom analyzer written in Cypher
 - c. Using a custom analyzer written in Java
3. **Hard:** How can you boost a specific field during a query ?
 - a. Using the ~ operator
 - b. Using the ^ operator
 - c. Using the + operator

Answer here: r.neo4j.com/hunger-games

Thank you

Christophe Willemsen
CTO - GraphAware
@ikwattro

graphaware.com



GraphAware®