

EXAMENSARBETE Categorization of Cypher Queries to Improve Benchmark Coverage for Graph Databases**STUDENTER** David Johansson, Jonathan Paul**HANDLEDARE** Simon Priisalu (Neo4j), Jens Wollert Ehlers (Neo4j), Per Runeson (LTH)**EXAMINATOR** Björn Regnell (LTH)

Testa mer effektivt med hjälp av användardata

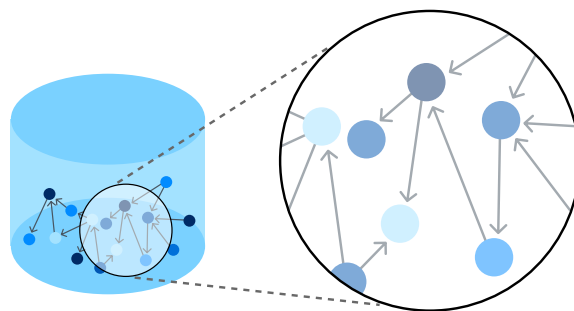
POPULÄRVETENSKAPLIG SAMMANFATTNING **David Johansson, Jonathan Paul**

Benchmarks är en form av tester som används för att mäta och upptäcka försämringar av kvaliteten i ett system. Dessa kan ofta vara uppbyggda med konstruerad användardata. I detta arbete har vi jämfört data från riktiga användare med benchmarks för att förbättra testerna för grafdatabasen utvecklad av Neo4j.

För att försäkra sig om att ett system bibehåller en hög kvalitet allt eftersom nya funktioner läggs till är benchmarks ett viktigt verktyg som hjälper utvecklare att se när problem i systemet dyker upp. Genom att upptäcka dessa problem innan de släpps till användare kan man förhindra att kvaliteten påverkas negativt. Användarupplevelsen är central för att bedöma kvaliteten och benchmarks behöver därför ta hänsyn till hur användaren nyttjar systemet. Ett sätt att göra detta på är att genom använda sig av användardata, som kan ge en bra indikation över vilka användningsområden som är vanliga i systemet.

Vi har i det här arbete jobbat tillsammans med det Malmöbaserade företaget Neo4j som tillhandahåller en grafdatabaslösning. En grafdatabas är uppbyggd av sammankopplade punkter av data, vilket illustreras i figur 1. För att en grafdatabas ska kunna hålla en hög kvalitet behöver den snabbt och effektivt hitta samband och relationer mellan olika punkter, vilket är anledningen till att benchmarks körs frekvent av utvecklarna på Neo4j. De existerande benchmarks som finns är skapade med hjälp av konstruerad användardata som exempelvis kan baseras på erfarenhet och utvecklarens kunskap av systemet. Neo4j samlar även in data från användare som vi i vårt arbete har jämfört med de redan existerande benchmarks som finns för grafdatabasen. Det gjorde vi genom att skapa ett verktyg som visar vilka funktioner

av databasen som saknar benchmarks. Med verktyget kan vi öka täckningsgraden så att vi inkluderar användarområden som inte omfattas av de benchmarks som finns idag. På så sätt kan Neo4j öka relevansen av sina benchmarks och se till att de testar den funktionalitet som används av deras kunder. Det här leder i sin tur att Neo4j kan bibehålla den höga kvaliteten som behövs för en snabb och effektiv databas.



Figur 1: Illustration av en grafdatabas¹.

Genom verktyget som vi utvecklat i arbetet kunde vi öka täckningsgraden av benchmarks med hela 32%, vilket genom fortsatt användning av verktyget kommer att öka ännu mer. Genom arbetet kunde även andra områden identifieras där en mer detaljerad datainsamling hade gett en bättre indikation över hur databasen används.