



FernUniversität Hagen

– Faculty of Mathematics and Computer
Science –

Different query languages for graph databases

Seminar „Discovering Big Data“

Chair of Databases and Information Systems

presented by

Daniel Langhann

Registration number: 3788687

Mentoring : Prof. Dr. Uta Störl

INTRODUCTION

1.1 RESEARCH TOPIC

Graph databases have developed into a powerful tool for modeling and querying complex relations in various data types. They are particularly suitable for dealing with linked data and complex network structures that essentially consist of nodes and their edges (relations). Examples of use are e.g:

- Relationships in social networks
- Recommendation applications
- Biological networks

In this paper, the topic of graph databases and selected query languages designed for graph databases will be presented and contextualized. In particular, the path from different query languages to a uniform ISO standard is pointed out.

1.2 MOTIVATION

The aim of this paper is to provide an overview of the different languages and to highlight the respective advantages and disadvantages, their capabilities, and limitations, as well as their specific use cases.

1.3 STATE OF RESEARCH

Graph databases are a comparatively young technology in the field of database systems. Over time, various query languages have been developed with specific focuses and corresponding strengths and weaknesses. A common ISO standard should help to keep the technology and its application more generic. In this paper, important proprietary languages are presented and differentiated from each other. In the further course, a overview of the different languages is created.

1.4 APPROACH

Basically, relevant documentation on the various query languages is analyzed and the key points are highlighted. To support this, selected languages are tried out in a test project and the corresponding learnings are integrated into the work.

STRUCTURE

After a general introduction to the topic of „Different query languages for graph databases “, an overview of the different query languages is given. Furthermore, it is shown how the different languages build on each other and which dependencies exist between the languages. Subsequently, selected query languages that are to be classified as relevant are reviewed, tested and differentiated from each other. The relevance of the languages is defined in terms of their popularity and their influence on the ISO standard for query languages for graph databases. After presenting the individual selected query languages, the path to a uniform ISO standard is described. In addition, the ISO standard is presented in a compact form. The research paper concludes with a discussion and a final conclusion.

- 1 Introduction
 - 1.1 Problem statement
 - 1.2 Motivation and objectives
 - 1.3 Approach
- 2 Different query languages and their concepts
 - 2.1 Cypher
 - 2.2 Gremlin
 - 2.3 SPARQL
 - 2.4 Comparison and differentiation of selected languages and concepts
- 3 The ISO-Standard for GQL
- 4 Summary
- 5 Outlook and Discussion
- 7 Conclusion

TIMELINE

After the topic was assigned at the end of March, the literature research was completed in mid-April. A preliminary structure and the exposé were created. After submitting the exposé at the end of April, the actual research paper will be started. The goal is to complete the paper by mid-June in order to have two weeks for final corrections and small changes before the final submission.

Tabelle 3.1: Task Schedule

Task	Deadline	Week
Selection and assignment of topics		15
Literature search		16
Literature review		17
Literature search		17
Planning a draft structure		17
Writing and submitting an exposé		17
Create structure		18
Write introduction		18
Submit peer review		18
Write main part with subchapters		21
Submit draft paper		21
Write conclusion		22
Correction reading		23
Check formatting		24
Check bibliography and references		26
Submit final draft		27