Graph Database on Medical Information for Integrated Life Science Research

Abstract

Indonesia is having an increasing surge of published scientific articles during recent years. In medical science, published articles greatly vary from both pre-clinical and clinical studies where each study possesses different methodological approach and hypothetical premises. However, some articles do not include a rigorous documentation as to make it reproducible. Moreover, the lack of centralized database further impedes researcher from re-analyzing previous findings and integrating them with the new study. This paper delineates such an issue by constructing a graph database to centralize and integrate medical information. Database is constructed using `Neo4j` and `cypher` querying language populated with 5,000 medical records generated by `synthea` program. Our database able to answer queries requiring complex relationship while minimizing the amount of db-hits, whereas such queries would require multiple joints should it be done using relational database. As a concluding remark, graph database is quite performant to solve data integration and centralization issue faced by life science research institutes.

1 Introduction

S: Increasing amount of published scientific article in Indonesia
P: Published paper is not highly reproducible
Q: How to tackle this issue?
R: Graph database to centralize and integrate research findings

1.1 Research in Indonesia

- Show data on how much Indonesia progresses within the last decade

1.2 Medical Informatics

- define information in medicine
- intricate nature on how one information relate with the other
- how graph database can help

1.3 Database

- define database
- type of database
- excerpt on positives and negatives
- graph db: implies underlying relationship as part of db
- relational db: treats relationship as joint of tables

2 Method

Graph database is constructed using `Neo4j` and `cypher` querying language.

- Create schema: constraints and indices
- Use periodic commit: free up memory when reaching certain threshold. Good for dealing with data with numerous entries
- Load file as a query object
- Set entity and relationship

3 Result

Constructed database is able to return answer to queries requiring with complex relationship, where otherwise in relational database it may require multiple joints to achieve the same answers.

- Show query example
- Show how many db-hits on a query
- Show figures describing query processing (require `Neo4j` browser)

4 Discussion

- Intro: reproducibility problem faced by researchers
- Body: graph db as a possible solution, considering the intricate relationships of medical information
- Premises on how graph db can be further utilized, maintained and developed
- Conclude the usability of graph db

5 Conclusion

Graph database is quite performant to integrate medical health record generated for 5,000 subjects using `synthea` program.

Reference