Graph Database on Medical Research Data 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 clinical research data. 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 database hits. As a concluding remark, graph database is quite performant to solve data integration and centralization issue faced by life science research institutes.