

Abstract

Graph database is a representation of modeling a collection of data into a form and Edge Node. Graph database is also a form or model of the database that provides effective and efficient solution for data storage. Developed in bigdata era like now is a new breakthrough in the field of Computer Science Engineering Data especially. Consisting of Edges, Nodes, and Properties are used to represent and store data. Are index-free adjacency, which means that each element contains a direct pointer to the adjacent elements and there is no search index is required. General graph database that can store any different graphs of specialized graphics such as triplestores database and network databases.

When only using the model database in the form of relational databases is certainly becoming increasingly difficult because the data here is very much at all. This is where the author will use the database model is still relatively new, ie Graph Database. This model can represent a lot of data in a graph that can be analyzed and conclusions drawn from the number of nodes and arcs that the authors obtained from molecular dataset chemical bonds. By using this model, of course, can be seen a summary of which can be seen from the molecular analysis and graph database summarization, authors take as the topic of the writing of this work. Summarization method that the authors take is RP-GD Algorithm that author use have an efficiency and quality can summarize a graph database. The algorithm is expected to improve the quality of a graph database so summarization of these models have maximum results in molecular represents the chemical bonding of the dataset.

From the test results and analysis, it is evident that the RP-GD algorithm can be used in summarization graph database, as well as produce a good quality in processing and results. Judging from the number of nodes and edges summarization results and coverage summarization ratios information and a parameter that indicates the results. Variations of summarization results can also be carried out in accordance with the desired minimum support. Value range of information from a summary graph database is proportional to the value of a given minimum support, while the ratio of summarization inversely proportional to the value of a given minimum support.

Keywords : *Graph Database, RP-GD Algorithm, dataset SMILES, summarization graph, summarization quality, chemical compounds, chemical informatics*