

## Abstrak

*Graph Database* merupakan representasi dari pemodelan suatu koleksi data ke dalam bentuk *Node* dan *Edge*. *Graph Database* adalah salah satu metode implementasi dari NoSQL (*Not Only SQL*), yaitu sistem database yang berguna untuk penyimpanan data dalam jumlah besar dan direpresentasikan dalam bentuk graph, sehingga data memiliki aksesibilitas yang tinggi. Namun data yang disimpan dalam pemrosesan *Graph Database* masih belum efisien dalam hal penyimpanan data. Penyimpanan jutaan ataupun milyaran *Nodes* dan *Edges* memerlukan pengompresan dalam kebutuhan penyimpanan data.

Dalam penelitian ini, kompresi *Graph Database* yang akan dilakukan adalah dengan menggunakan GRAC (*Graph Algorithm Clustering*). *Graph Database* yang digunakan yaitu suatu data yang berisikan data kolaborasi antar penulis jurnal ilmiah. Dalam GRAC (*Graph Algorithm Clustering*), Clustering yang digunakan adalah *Hierarchical Clustering*. Metode *Hierarchical Clustering* adalah suatu metode dalam Clustering yang akan mengcluster *Node* menjadi bentuk *Cluster Node* secara hirarki. Dalam pembuatan cluster yang hirarki, strategi yang dipakai adalah *Agglomerative* dimana setiap *Node* nantinya akan digabungkan menjadi satu cluster. Untuk mendapatkan strategi *Agglomerative* yang efektif dan efisien maka akan dihitung jarak maximum antar cluster yang biasa disebut *Complete Linkage Clustering*. Setiap *Node* terlebih dahulu dihitung *Jaccard index*nya untuk mendapatkan bobot jarak antar *Node*. Penggunaan *Hierarchical Clustering* adalah untuk membentuk *Cluster Node* yang memiliki kesamaan tetangga. *Cluster Node* nantinya akan dihubungkan dengan *Cluster Edge* dimana, *Cluster Edge* didapatkan melalui pencarian secara greedy pada setiap hubungan *Cluster Node* yang mengabstaksi *Edge* paling banyak.

Dengan menerapkan GRAC (*Graph Algorithm Clustering*) dengan menggunakan metode *Hierarchical Clustering* yang membentuk cluster yang hirarki, maka akan menghasilkan graph database yang bersifat *lossless* serta terkompres dengan baik.

**Kata Kunci:** *graph database, compression, graph algorithm clustering, Hierarchical Clustering, agglomerative, complete linkage clustering, jaccard index, greedy, scientific journal authors*

## Abstract

Graph Database is a representation of data collection that formed in to Node and Edge. Graph Database is a one of implementation method in NoSQL (Not Only SQL), database system that can store big data and represent in the form of graph, that having a great performance. But, the data that store in the process of Graph Database is not efficient in that way. Storing milion of Node and Edge need compression to store the data.

In this research, Graph Database Compression using the GRAC(Graph Algorithm Clustering) method. Graph Database use the data that carrying the collaboration of scientific journal author. In the GRAC(Graph Algorithm Clustering) method, Clustering using Hierarchical Clustering. Hierarchical Clustering method is the method that cluster Node in to Cluster Node Hierarchically. The form of Hierarchical Clustering, strategy that uses is Agglomerative that every node merge into one Cluster. Complete Linkage Clustering is using in the Agglomerative strategy for count the maximum distance of Cluster. Every Node is counted the jaccard index for the distance index. Hierarchical Clustering is used to create the Cluster Node that have the same neighbourhood. Cluster Node then connect to the Cluster Edge that Cluster Edge known from greedy searching in the relationship of Cluster Node that having the most neighbourhood.

Using the GRAC (Graph Algorithm Clustering) with Hierarchical Clustering method that create cluster Hierarchically, then create the graph database Lossless.

**Kata Kunci:** *graph database, compression, graph algorithm clustering, Hierarchical Clustering, agglomerative, complete linkage clustering, jaccard index, greedy, scientific journal authors*