

Abstrak

Basisdata graf merupakan representasi dari pemodelan suatu koleksi data ke dalam bentuk *Node* dan *Edge*. Basisdata Graf juga merupakan sebuah bentuk atau model dari *database* yang menyediakan solusi efektif dan efisien terhadap penyimpanan data. Dikembangkan di era *bigdata* seperti sekarang ini merupakan terobosan baru di bidang *Computer Science* khususnya *Data Engineering*. Terdiri dari *Edges*, *Nodes*, dan *Properties* yang digunakan untuk merepresentasikan dan menyimpan data. Bersifat *index-free adjacency* yang berarti bahwa setiap elemen berisi pointer langsung ke elemen yang berdekatan dan tidak ada pencarian indeks diperlukan. *Database* grafik umum yang dapat menyimpan grafik pun berbeda dari *database* grafik khusus seperti *triplestores* dan *database* jaringan.

Ketika hanya menggunakan model *database* yang berbentuk *relational database* tentunya semakin lama semakin kesulitan karena datanya disini sangatlah banyak sekali. Disinilah penulis akan menggunakan model *database* yang masih tergolong baru, yaitu *Graph Database*. Model ini dapat merepresentasikan banyak data dalam suatu graf yang bisa dianalisis serta diambil kesimpulannya dari banyak simpul serta busur yang penulis peroleh dari dataset molekuler ikatan kimia. Dengan menggunakan model ini, tentunya dapat dilihat ringkasan molekuler yang dapat dilihat dari analisa dan peringkasan basisdata graf yang penulis ambil sebagai topik dari penulisan karya ini. Metode peringkasan yang penulis ambil adalah *RP-GD Algorithm* yang penulis gunakan mempunyai efisiensi dan kualitas yang dapat meringkas suatu basisdata graf. Diharapkan algoritma tersebut bisa meningkatkan kualitas dari sebuah *graph database* sehingga peringkasan dari model tersebut mempunyai hasil yang maksimal dalam merepresentasikan molekuler ikatan kimia dari dataset tersebut.

Dari hasil pengujian dan analisis, maka terbukti bahwa algoritma *RP-GD* dapat digunakan dalam peringkasan basisdata graf, serta menghasilkan kualitas yang baik dalam pemrosesan maupun hasilnya. Dilihat dari jumlah nodes dan edges hasil peringkasan lalu cakupan informasi dan rasio peringkasan menjadi parameter yang menunjukkan hasil tersebut. Variasi hasil peringkasan juga dapat dilakukan sesuai dengan minimum support yang diinginkan. Nilai cakupan informasi dari sebuah ringkasan basisdata graf berbanding lurus dengan nilai *minimum support* yang diberikan, sedangkan rasio peringkasan berbanding terbalik dengan nilai *minimum support* yang diberikan.

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