

ABSTRAK

Penggunaan *graph* dalam memodelkan suatu struktur yang rumit saat ini berkembang secara pesat terutama dalam memodelkan struktur seperti susunan molekul, jaringan protein, dan jaringan sosial. Penggunaan *graph database* untuk menangani tipe data *graph* yang memiliki relasi yang kompleks dinilai lebih efektif daripada menggunakan *relational database*.

Dalam mempercepat pemrosesan *query* pada *graph database* dibutuhkan suatu metode yang dapat disebut *graph indexing* agar lebih cepat dan efisien. GIndex merupakan salah satu metode *graph indexing* yang mendukung pemrosesan *query* bertipe *subgraph query*.

Pada metode GIndex menerapkan beberapa teknik seperti *size-increasing support constraint* untuk membangun *feature set database* dan pemilihan *discriminative fragments* dalam membangun *index*. Kemudian membandingkan data pada *index* dengan *feature set query* untuk mendapatkan *candidate set* yang nantinya akan dilakukan *subgraph matching* menggunakan algoritma Ullman untuk mendapatkan *answer set*.

Pada pengerjaan tugas akhir ini data yang akan dijadikan sebagai *dataset* merupakan susunan molekul. Berdasarkan pengerjaan tugas akhir yang telah dilakukan didapatkan hasil bahwa pada pengimplementasian algoritma GIndex, jika menggunakan nilai *maximal frequent fragment* yang cukup besar maka akan memakan waktu yang lebih lama dan memungkinkan jumlah *candidate set* yang didapatkan akan lebih sedikit, berlawanan dengan penggunaan nilai *minimal discriminative fragment*. Banyaknya jumlah *candidate set* yang didapatkan akan berpengaruh pada waktu *subgraph matching* yang dibutuhkan.

Kata kunci: *graph, graph database, GIndex, subgraph query, size-increasing support constraint, discriminative fragments, index, subgraph matching*