

## Abstrak

Kebanyakan *search engine* saat ini menggunakan sistem perengkingan pada dokumen hasil pencarian. Sehingga diperoleh begitu banyak dokumen yang sebenarnya tidak semuanya relevan menurut user. Akhirnya user harus membuka satu per satu dokumen untuk memastikan apakah dokumen tersebut yang user inginkan. Oleh karena itu pada tugas akhir ini penulis menggunakan metode *clustering* pada dokumen-dokumen hasil pencarian untuk mengatasi permasalahan tersebut.

Metode *clustering* yang digunakan adalah metode *Generalized Incremental (GenIc)*. GenIc menggunakan teknik *Generalized incremental* dan *evolutionary* dalam melakukan proses pengelompokan dokumennya. Yang dimaksud dengan *Generalized incremental* adalah proses penggeseran *center* sehingga jarak *center* dengan dokumen pada cluster tersebut semakin dekat. Sedangkan, teknik *evolutionary* adalah proses update *center* dengan menghitung probabilitas tiap *center* sehingga dapat ditentukan mana *center* yang dapat “bertahan” atau “dibunuh”. Untuk menerapkan metode ini, ada beberapa parameter yang harus dimasukkan oleh user yakni jumlah cluster awal( $m$ ), jumlah cluster akhir( $k$ ), jumlah generasi( $n$ ), dan nilai  $\delta$ .

Pengujian dilakukan untuk menganalisis pengaruh keempat parameter terhadap performansi waktu GenIc dan kualitas cluster yang dihasilkan. Kualitas cluster dilihat dari nilai *similarity* antar dokumen dalam satu cluster. Semakin tinggi nilai *similarity*-nya maka semakin bagus kualitas cluster yang dihasilkan dan demikian sebaliknya.

**Kata kunci:** *clustering*, algoritma GenIc, teknik *evolutionary*, *generalized incremental*