ABSTRAK

Email sudah menjadi alat penting untuk berkomunikasi dalam kehidupan sehari-hari. Tinggi nya jumlah *email* yang masuk dapat mempersulit pengguna dalam mengelompokkan *email* secara manual. Melalui proses *clustering* dengan metode K-Means, X-Means Heuristic, dan X-Means Dynamic dibantu dengan fitur pembobotan *Ranking Term Frequency* dan TF-IDF, data *email* dapat diolah agar dapat dikelompokkan secara otomatis. *Cluster* yang terbentuk juga akan diberi label secara otomatis, agar setiap *cluster* memiliki label berdasarkan isi dari data *email* yang sudah di clustering. Sehingga perlu dilakukan analisis, perancangan, desain aplikasi, dan pencarian teori-teori yang mendukung untuk membuat aplikasi email clustering dengan menggunakan algoritma k-means, x-means heuristic, dan x-means dynamic. Pengujian yang dilakukan pada 20 akun email yang berbeda, ditarik 100 data *email* lalu diuji dengan 3 tahapan berbeda dan 3 metode *cluster*. Hal ini dilakukan untuk mengimplementasikan aplikasi *email clustering* agar dapat diuji tingkat keakuratan nya dalam melakukan proses *email clustering*. Aplikasi *email clustering* ini dapat mengelompokkan *email* dan memberikan label secara otomatis dengan rata-rata akurasi sebesar 90,08%, menggunakan penarikan data *subject* dan *body email*.

Kata Kunci: Algoritma K-Means, Algoritma X-Means Heuristic, Algoritma X-Means Dynamic, *Ranking Term Frequency*, TF-IDF

ABSTRACT

Email has become an important tool for communicating in everyday life. The high number of incoming email may complicate the user to classify the email manually. Through the process of clustering with K-Means method, X-Means Heuristics, and X-Means Dynamic, augmented with Ranking Term Frequency weighting features and TF-IDF,emails data could be grouped automatically. Clusters that formed also be labeled automatically, so that each cluster has a label based on the contents of the email data that is already in cluster. So that needs to be done the analysis, design, application design, and search the theories that support to make the email application clustering using k-means algorithm, x-means heuristic, and x-means dynamic. Tests were performed on 20 different email accounts, then downloaded 100 emails data, and last tested with 3 different stages and three methods of cluster. This was done to implement email clustering application so that it can be tested its accuracy in the process of email clustering. This application can classify emails and give labels automatically with an average accuracy of 90.08%, using downloaded data subject and body emails.

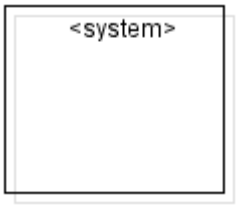
Keywords: K-Means Algorithm, X-Means Heuristic Algorithm, X-Means Dynamic Algorithm, Ranking Term Frequency, TF-IDF

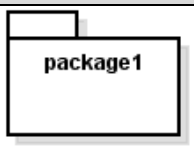
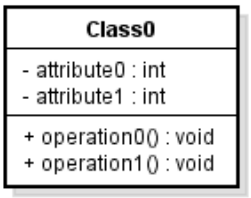
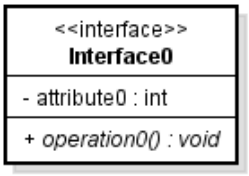
DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR NOTASI/ LAMBANG

Use Case Diagram (UML 2.0)		
No	Gambar	Keterangan
1.		Menggambarkan aktor atau pengguna aplikasi.
2.		Menggambarkan proses atau aksi yang dapat dilakukan oleh aktor pada aplikasi.
3.		Menggambarkan sistem tempat proses dijalankan

Class Diagram (UML 2.0)		
No	Gambar	Keterangan
1.		Menggambarkan paket tempat menyimpan sekumpulan kelas
2.		Menggambarkan sebuah kelas beserta atribut dan *method*-nya
3.		Menggambarkan sebuah *interface* beserta atribut dan *method*-nya

Activity Diagram (UML 1.3)		
No	Gambar	Keterangan
1.		Menandakan dimulainya aktivitas pada sebuah sistem.
2.		Menandakan aktivitas apa yang akan dilakukan oleh pengguna aplikasi.
3.		Menandakan akhir aliran proses sistem