

ABSTRAK

DISTRIBUTED TWITTER CRAWLER

Twitter sebagai sebuah situs jejaring sosial memberikan akses kepada penggunanya untuk mengirimkan pesan singkat maksimal 140 karakter yang disebut dengan *tweet*. *Tweet* dapat pula berupa gambar. Berbagai macam manfaat dapat diperoleh dari *tweet* misalnya *event detection*, prediksi pergerakan saham, prediksi pemilu, penyebaran penyakit, dan sebagainya. Ide dasar pembuatan penelitian ini adalah keterbatasan untuk mendapatkan *tweet* jika hanya mengandalkan satu buah pengumpul *tweet*. Penelitian ini difokuskan pada pengembangan *prototype* untuk memproses pengumpulan (*crawler*) *tweet* yang terdistribusi. Pemanfaatan sistem terdistribusi digunakan untuk menangani keterbatasan permintaan pada Twitter API. Pada proses pendistribusian, kata kunci pengumpulan *tweet* diolah oleh *node* yang berperan sebagai koordinator terhadap *node* lain yang ada. Jika *node* koordinator mengalami gangguan (*error*), secara otomatis *node* lain akan melakukan pemilihan koordinator yang baru dengan menggunakan algoritma ring. Hasil uji coba menunjukkan bahwa aplikasi dapat mendistribusikan proses pengumpulan *tweet* ke dalam tiga buah *node*. Luaran dari aplikasi ini dapat berguna bagi para analis yang ingin mengolah data yang bersumber dari jejaring sosial khususnya twitter.

Kata kunci: *Crawler*, Twitter, Sistem Terdistribusi, Algoritma Ring

ABSTRACT
DISTRIBUTED TWITTER CRAWLER

Twitter as a social networking site provides access to users to send short messages up to 140 characters called tweets. Tweets can also be pictures. Various kinds of benefits can be obtained from the tweet such event detection, prediction of stock movement, election prediction, the spread of disease, and so on. The basic idea of making this study is limited to get a tweet if only rely on one crawler of tweet. This study is focused on the development of a prototype for distributed processing tweet collection (crawler). Utilization of a distributed system is used to handle the limitations request on Twitter API. In the process of distributing, keywords of collecting tweets processed by the node that acts as the coordinator of the other existing nodes. If the coordinator node get impaired (error), other nodes will automatically elect a new coordinator by using algorithms ring. Experimental results show that the application can distribute the process of collecting tweets into three nodes. Outcomes of these applications can be useful for analysts who want to process data from social networks, especially Twitter.

Keywords: Crawler, Twitter, Distributed System, Ring Algorithm