

Pendeteksian topik pada twitter menggunakan online eigenspace-based fuzzy c-means clustering untuk big data = Topic detection on twitter using online eigenspace-based fuzzy c-means clustering for big data / Christhoper Nugraha

Christhoper Nugraha, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493866&lokasi=lokal>

Abstrak

ABSTRAK

Deteksi topik adalah proses menganalisis kumpulan data tekstual untuk menentukan topik pengumpulan data tekstual. Salah satu metode pengelompokan yang dapat digunakan untuk deteksi topik adalah metode Fuzzy C-Means (FCM). Namun, penggunaan FCM sederhana untuk pendeteksian topik tentang big data kurang efektif, karena akan memakan waktu lama dan banyak memori. FCM sederhana juga memiliki masalah lain, ketika melakukan deteksi topik aktif data dimensi tinggi, FCM sederhana hanya akan menghasilkan satu topik. Dalam penelitian ini, suatu gabungan metode Single-Pass Fuzzy C-Means (SPFCM) dan Fuzzy C-Means Berbasis Eigenspace (EFCM) diusulkan, yaitu Single-Pass Eigenspace-Based Fuzzy C-Means (SPEFCM) metode untuk mengatasi masalah ini. Data yang digunakan untuk deteksi topik adalah

tweet yang berasal dari aplikasi Twitter. Lalu, keakuratan topik didapat menggunakan SPEFCM dan EFCM akan dibandingkan berdasarkan nilai koherensi. Itu hasil simulasi menunjukkan bahwa nilai koherensi topik yang diperoleh menggunakan SPEFCM adalah sebanding dengan EFCM. Ini menunjukkan bahwa SPEFCM adalah metode yang tepat untuk mendeteksi topik pada data besar, tanpa mengurangi kualitas topik yang dihasilkan.

ABSTRACT

Topic detection is the process of analyzing a textual data set to determine the topic of textual data collection. One of the grouping methods that can be used for topic detection is the Fuzzy C-Means (FCM) method. However, the use of simple FCM for the detection of topics about big data is less effective, because it will take a long time and a lot of memory. Simple FCM also has another problem, when detecting active topics of high dimensional data, simple FCM will only produce one topic. In this study, a combination of the Single-Pass Fuzzy C-Means (SPFCM) method and the Fuzzy C-Means Based on Eigenspace (EFCM) is proposed, namely the Single-Pass Eigenspace-Based Fuzzy C-Means (SPEFCM) method to overcome this problem. The data used for topic detection is tweets that come from the Twitter application. Then, the accuracy of the topics obtained using SPEFCM and EFCM will be compared based on coherence values. The simulation results show that the topic coherence value obtained using SPEFCM is comparable to EFCM. This shows that SPEFCM is the right method for detecting topics in big data, without reducing the quality of the topics produced.