

Analisis kinerja sistem klasifikasi video berbasis tag menggunakan mapreduce untuk internet content profiling = An performance analysis of video classification system based on tag using mapreduce for internet content profiling

M. Adhika Putra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20430188&lokasi=lokal>

Abstrak

Sebagian besar informasi yang beredar di internet merupakan konten video. Informasi video ini perlu dianalisis karena tidak semuanya yang beredar adalah video dengan konten yang baik. Banyak video dengan konten yang buruk beredar luas di internet dan dapat diakses oleh siapapun yang mengakses internet. Pada penelitian ini, dibuat sistem klasifikasi video pada Youtube dengan metode Symbolic Distance dan Focal Point menggunakan model pemrograman MapReduce pada Hadoop. Sistem klasifikasi ini mengidentifikasi tag yang tersemat pada setiap video di Youtube kemudian dibandingkan dengan matriks co-occurrence untuk mencari nilai symbolic distance pada sebuah video. Penggunaan metode Focal Point pada sistem klasifikasi bertujuan untuk meningkatkan akurasi dan fokus untuk klasifikasi video.

Dalam penelitian ini diukur juga kecepatan pemrosesan sistem klasifikasi dengan menggunakan Hadoop serta dicari faktor-faktor yang dapat mempengaruhi kecepatan pemrosesan. Untuk itu dilakukan 3 skenario pengujian berdasarkan ukuran InputSplit yang digunakan, jumlah node, serta konfigurasi pada YARN masing-masing dengan 3 ukuran file (500 MB, 1 GB, 1,5 GB) dengan masing-masing jumlah tag sebesar 58718, 119697, dan 160395 tag. Pada file berukuran 500 MB, 1 GB, 1,5 GB, penambahan jumlah node dapat mempercepat kecepatan rata-rata pemrosesan sebesar 0,2 detik, 5 detik, dan 16,3 detik. Kemudian dengan melakukan konfigurasi pada YARN, kecepatan pemrosesan dapat dipercepat hingga 47 detik, 277,1 detik, dan 354,3 detik pada file berukuran 500 MB, 1 GB, 1,5 GB. Dari pengujian juga diketahui semakin kecil InputSplit maka semakin tinggi kecepatan pemrosesan MapReduce. Namun jika mapper tidak dapat menangani jumlah split yang ada, maka kecepatan pemrosesan data akan menjadi lebih lambat dari sebelumnya.

.....Most information that widely spread on the internet is video. This video information needs to be analyze because not all of the information have a good content. There are many video with bad content widely spread on the internet and anyone can access that video easily. In this research, Youtube Video Classification System with Symbolic Distance and Focal Point Method is made using a MapReduce from Hadoop framework. This system identifying the tag that assign in every Youtube video and then compare the tag with co-occurrence matrix to find the symbolic distance value for a single video. Focal Point in this system is useful to improve accuracy and focus of video classification.

This research will measure the processing speed of this classification system and then search the factor that can affect processing speed. For that, three skenarios are implemented based on InputSplit size, amount of node, and YARN configuration with three file size (500 MB, 1 GB, 1,5 GB) with the number of each tag are 58718, 119697, and 160395 tag. For file with 500 MB, 1 GB, and 1,5 GB size, increasing the amount of node from two to three can speed up the process for 0,2 second, 5 second, and 16,3 second. Optimize the YARN configuration can speed up the process for 47 second, 277, 1 second, and 354,3 second for file with size of 500 MB, 1 GB, and 1,5 GB. This Reasearch also discover that if the size of InputSplit is small, then

the speed of data processing is faster. But if the mapper can't handle the amount of the split, it can make the processing speed slower than before.