

Abstrak

Halaman web spam menggunakan berbagai macam teknik untuk mencapai peringkat tinggi pada mesin pencari. Manusia dapat melakukan pekerjaan dengan baik untuk mengidentifikasi *web spam* tersebut, tetapi untuk mengetahui web spam secara praktis tidak layak menggunakan usaha manusia untuk mengetahui sejumlah besar halaman. Mirip dengan algoritma *TrustRank*^[2], Algoritma *Anti-TrustRank* merupakan sebuah metode perangkian web yang berusaha untuk mengatasi *link based spam* dengan mengusung ide bahwa *web spam* biasanya akan membentuk *link farm* yang berisi spam-spam lain, dalam artian *web spam* biasanya akan merefernsikan ke *web spam* lain. Cara yang diusulkan metode ini adalah dengan membentuk *seed set* yang berisi beberapa *web* yang telah diketahui secara pasti statusnya sebagai *web spam* dan dipastikan mempunyai nilai PageRank yang tinggi. Kemudian dari *seed set* nilai *antitrust* didistribusikan ke seluruh *web* di luar *seed set*. Jumlah *seed set* mempengaruhi hasil dari Algoritma *Anti TrustRank*. Jumlah *seed set* yang lebih banyak akan mendapatkan presisi yang lebih baik dalam mendapatkan *web spam* yang berada diluar *seed set*. Selain *seed set*, jumlah iterasi juga sangat berperan dalam menentukan hasil dari algoritma *Anti TrustRank*, karena jumlah iterasi merupakan parameter yang digunakan dalam menentukan jarak antar *web* dalam graf. Pada percobaan dalam dataset WebGraph^[12] tersebut menunjukkan bahwa pendekatan *Anti-TrustRank* efektif dalam mendeteksi halaman *spam* walaupun dengan jumlah *seed set* yang kecil jika dibandingkan dengan algoritma *TrustRank*.

Kata kunci: *Anti-TrustRank, TrustRank, seed set, web spam.*