

Abstrak

Identifikasi web spam merupakan suatu tantangan yang besar bagi *search engine*. *Web spam* dengan sengaja melakukan manipulasi posisinya dengan menarik perhatian *surfer* dalam proses perankingan pada *search engine*. Salah satu teknik yang digunakan oleh *spammer* yakni *link farm* yang merupakan kumpulan *link* web yang saling terhubung untuk mendapatkan *PageRank* yang tinggi. Oleh karena itu apabila diketahui salah satu *web spam* maka *link farm* dapat diidentifikasi, dan dapat menganggap *web* yang terhubung dalam *link farm* tersebut adalah *web spam*.

BadRank adalah suatu metode yang digunakan untuk mendeteksi *web spam* berdasarkan anggapan bahwa suatu *web* akan dianggap *spam* apabila memiliki keterhubungan *link* dengan *web spam*. Metode *BadRank* perlu dilakukan modifikasi untuk memastikan nilai *BadRank* yang diperoleh konvergen dengan memastikan matriks yang terbentuk stokastik. Selain itu metode ini juga dapat dioptimasi dengan cara menambahkan variabel yang berisi web yang diyakini bukan *spam (trust)*. Dalam tugas akhir ini digunakan dataset WEBSHAM-UK 2006 untuk pengujian *BadRank* dengan modifikasi stokastik dan *trust*.

Dari hasil pengujian diperoleh bahwa untuk dataset *web spam uk 2006*, metode *badrank* dengan modifikasi *leafbadlinks with trust* mampu mendeteksi *spam* yang lebih baik 1%-3% dibandingkan dengan modifikasi lainnya. Selain itu juga penambahan variabel *trust* mampu untuk lebih mengoptimalkan 3% -5% deteksi *web spam* pada metode *badrank*.

Kata kunci: *Web Spam, Badrank, Link Farm*