

Abstrak

Sebagian besar dari informasi pada Web diisi di struktur obyek yang teratur, disebut dengan *data record*. *Data record* ini sangat penting karena mempresentasikan inti informasi dari halaman host-nya, misalnya daftar produk atau layanan. *Me-mining data record* untuk mengekstrak informasi dari halaman Web bertujuan untuk menyediakan nilai tambah suatu layanan.

Pada Tugas Akhir ini diimplementasikan metode untuk *mining data record* pada halaman Web secara otomatis dengan menggunakan algoritma yang disebut MDR (*Mining Data Records in Web Page*). Teknik ini lebih efektif karena hanya berdasarkan pada dua pengamatan penting saja, yaitu mengamati *data record* yang berada pada halaman Web dan algoritma pencocokan string. Pada proses *me-mining data record* ini ada tiga langkah yang utama yaitu, membangun sebuah *tag tree HTML* dari halaman Web, *mining data region* pada halaman Web dengan menggunakan tag tree dan pencocokan string, dan mengidentifikasi *data record* dari setiap *data region*.

Tahap analisis dan pengujian memberikan hasil bahwa algoritma MDR yang dibangun terbukti bisa mendapatkan *data record* pada halaman Web meskipun ada beberapa *noise*.

Kata kunci: Web Mining, tag tree HTML, data region, data record.