

Abstrak

Website memiliki banyak konten yang ditampilkan berdasarkan blok-blok yang ada dalam halaman web. Dalam halaman web yang ditampilkan tidak semuanya merupakan inti informasi dari halaman web tersebut misalnya iklan, panel navigasi, *user guide*, *links* dan sebagainya. Data-data penyusun informasi yang ditampilkan dalam sebuah web akan dibentuk dalam sebuah struktur obyek yang teratur dan memiliki kerangka baku yang akan kita kenal sebagai *data record*. *Data record* ini yang akan dijadikan sebagai pendekatan untuk mengekstrak informasi dalam halaman website, dalam bahasa lain dapat kita katakan sebagai proses mining *data record*. Hal ini dilakukan dengan tujuan untuk menyediakan nilai tambah suatu layanan website.

Pada Tugas Akhir ini diimplementasikan metode untuk *mining data record* pada halaman Web secara otomatis dengan menggunakan algoritma yang disebut DEPTA (*Data Extraction with Partial Tree Alignment*)[1]. Teknik ini dibuat berdasarkan pengembangan dari algoritma MDR (*Mining Data Records in Web*)[2] yaitu dengan menambah tiga tahap, diantaranya adalah *tree edit distance* yang diimplementasikan bersama algoritma *simple tree matching* dan yang terakhir adalah *partial tree alignment* dalam tiga tahap tersebut selalu menggunakan prinsip pencocokan dua atau lebih *subtree*.

Tahap analisis dan pengujian memberikan hasil bahwa algoritma DEPTA yang dibangun terbukti bisa mendapatkan *data record* pada halaman Web dan mampu mengurangi tingkat *noise* pada hasil akhir yang lebih baik dari pada algoritma MDR.

Kata kunci: Web Mining, partial tree alignment, data region, data record, simple tree matching, mining data record.