Abstrak

Web di internet telah menjadi *repository* data yang luar biasa besarnya. Telah banyak upaya yang dilakukan untuk menyediakan akses yang efisien terhadap informasi yang relevan didalam *repository* data yang sangat besar ini. Salah satu cara untuk menyediakan akses yang efisien ini adalah dengan cara *web news content extraction* yang memiliki fokus utama mengambil informasi dalam web berita.

Pada Tugas Akhir ini diimplementasikan metode untuk mengekstrak informasi utama pada halaman Web berita dengan menggunakan metode yang disebut *Hybrid*. Teknik ini berusaha mengambil keuntungan dari teknik *sequence matching* dan *tree matching*. Struktur data yang digunakan adalah TSReC, yang merupakan salah satu representasi *tag sequence* yang sesuai untuk kedua teknik *sequence matching* dan *tree matching*.

Tahap analisis dan pengujian memberikan hasil bahwa metode *Hybrid* yang dibangun terbukti bisa mendapatkan *news content* pada halaman Web berita meskipun pada beberapa dataset masih terdapat *noise*.

Kata Kunci : *web news content extraction, sequence matching, tree matching,* TSReC.