

Abstrak

Penggunaan Internet semakin pesat dan orang membutuhkan suatu cara untuk melihat content-content yang penting dari sebuah halaman Web. Hal inilah yang mendorong diciptakannya suatu teknologi untuk mengekstrak *content* atau informasi dari halaman Web tersebut sehingga dapat mempermudah dalam pembacaan dan analisis. Informasi pada halaman Web dapat berupa teks, gambar, alamat URL dan sebagainya. Karena bentuknya yang *semi-structured*, untuk mengambil informasi dari halaman Web cukup sulit.

Wrapper merupakan salah satu metode untuk mengekstrak halaman Web. Namun *Wrapper* mempunyai kelemahan, yaitu tidak adanya proses *learning*, sehingga sistem berjalan secara manual (*hand coded*), karena itulah dibuat suatu metode pengembangan dari *Wrapper* ini yang menyediakan proses *learning* yaitu *Wrapper Induction*. Proses *learning* pada *Wrapper Induction* ini adalah pada proses generate tag *HTML* sebagai penentu content-content yang akan diekstrak. Pada Tugas Akhir ini akan dilakukan ekstraksi informasi yang berupa teks berita menggunakan *Wrapper Induction* dan analisis perfomansi dari *Wrapper Induction* dalam mengekstrak halaman web berdasarkan *Recall*, *Precision* dan *F-Measure*.

Kata kunci : *Wrapper*, *Wrapper Induction*, halaman Web