

Abstract

The using of Internet is increase and people need a technique to get the important contents of a Web page. Because of that case, the technology to extract contents or information of a Web page had been invented, then the Web page can be both read and analyzed easily. Web page contains of many informations such as text, images, URL address and so on. Because of semi-structured, there's quite difficult to take information from Web Page.

Wrapper is a one of methods to extract a web page. But, Wrapper has a weakness; it doesn't have a learning process, then the system running manually (hand coded). Because of that case, Wrapper Induction which is provided a learning process had developed. Learning process on Wrapper Induction is a process to generate HTML tag to indentify which content will be extract. This Final Project is created to extract text information from news Web page using Wrapper Induction and analyze the performance of Wrapper Induction on extracting a Web page based on Recall, Precision and F-Measure.

Keywords: Wrapper, Wrapper Induction, Web page