

Abstraction of Dissertation: (18/1/2012)

## **New Method for Building a Mashup Through DOM Tree Modeling System That Automatically Generated Using Xtractorz Algorithm**

**Dr. Ir. Rudy Agus Gemilang Gultom, M.Sc.**

Department of Electrical Engineering

Faculty of Engineering

The University of Indonesia

In this dissertation, we proposed a new method for building a Mashup through DOM tree modeling system that automatically generated by using Xtractorz algorithm. Mashup is a web based application which combining data or function from two external sources (web pages) or more in order to create new services in the new web pages. DOM tree is a useful platform to represent or to interact various HTML or XML web pages document object in a form of tree structure or commonly called node-tree. With the use of DOM tree structure approach then the web browser could interpret the HTML tags structure from a web page very easily. This mean that all the nodes exist inside of HTML tags structure within a web page can be easily identified or modified its data structures.

The aim of this new method is to ease the Internet users to build a Mashup via online Internet. In order to implement this new method, a new Mashup builder tool called Xtractorz has been proposed and build in this research project. The Xtractorz tool is consist of a main algorithm and its derivative algorithms completed with a set of rules which is designed to be capable to build a Mashup through a process of data table extraction from an HTML web page on the Internet. The process of building a Mashup is initiated by mapping the whole structure of a webpage HTML codes or tags that has been succesfully extracted and grouped from a single web page. Afterward, those HTML codes or tags that represent web HTML tables will be grouped into Root, Parent, Child, Sibling and Leaf Nodes by the Xtractorz main algorithm. Those group of nodes will be transformed to a DOM tree structure model that automatically build by Xtractorz tool. This DOM tree structure will be used as a main reference for computation process in every stages of building a Mashup, such as: Data Retrieval, Data (Source) Modeling, Data Cleaning/ Filtering, Data Integration and Data Visualization.

The end result of building a Mashup stages will be the variety of output as expected or needed by the users, for example in the form of PDF and XML tables and many more output forms. The new method of building a Mashup has been tested by testing the Xtractorz tool using the same scenario and environment test that has been done previously by similar Mashup builder tools, called: Karma and Dapper. The tests has been done by the appointed testers with the qualification of programmers and non-programmers background as representation of Internet users. The serial tests have been deployed with specific tasks model, namely Task 1, Task 2 and Task 3. Basically, the appointed testers created a Mashup as written on the scenario. The stage of building a Mashup is initiated by an on-line and real-time process of web data table extraction from the Internet.

The data results showed that the Xtractorz tool performance is better than the other tools, Karma and Dapper. This can be proofed by the successful test of Xtractorz tool which can completed the whole process of building a Mashup with minimum time spent or more faster. Beside that, the Xtractorz main algorithm and its derivatives can also completed the five stages standardization of building a Mashup, starting from Data Retrieval, Data Modeling, Data Cleaning/ Filtering, Data Integration until Data Visualization stage which can not be performed completely by other similar tools. As conclusion, the result test showed that this new method for building a Mashup through DOM tree modeling system which automatically generated using Xtractorz algorithm has a better performance that could help the Internet users to build a Mashup more faster and easily.

**Keywords:** Building a Mashup, DOM Tree, Web Table Extraction, Xtractorz Algorithm.

