

# IMPROVING WEB APPLICATIONS TESTING BY CLUSTERING USER SESSION DATA

Hsu Mon Maung  
University of Computer Studies, Mandalay  
hsumon77@gmail.com

## ABSTRACT

*Nowadays, web applications are used in various situations such as search, on-line shopping and on-line banking. Effective testing of web applications is crucial to provide reliable services for the fast growing demand on web application. Web log files provide a useful way for web application test. User session data are extracted from web log files and then created the test data. This paper describes the effective and efficient way of generating the test data from the large set of user session data. The clustering algorithm (k-means-like) is adopted as a reduction mechanism for partition user session data into a set of cluster. Samples of each cluster are selected and constructed into the test data. The metrics of clustering quality, dissimilarity definition of user sessions and code coverage with the generated test data are defined in this paper.*

Keywords: user session, clustering algorithm, dissimilarity function.

## 1. INTRODUCTION

Web applications have become critical components of the global information infrastructure, and it is important that they be validated to ensure their reliability. Web applications are among the fastest growing classes of software system in use today [1]. A plenty of approaches of testing web applications are proposed in literature [3]. Web application can be considered as distributed software with client-server structure [2]. The user at the client sends a request with a browser to the

server over the internet. The web server responds to the client by delivering the web pages. The web server records each interaction with user in log files. Log file is the view of client-server transactions which generally consists of client request with the server response. Multiple transactions can be required for a web page. Each transaction is totally separated in log files. There are standardized log formats: Common Log Format (CLF), Extended Common Log Format (ECLF), and W3C standard. Common Log Format of Apache web server looks like the followings:

```
111.111.111.111 - - [08/Oct/2007:11:17:55 -0400]  
"GET /style.css HTTP/1.1" 200 3225  
"http://www.loganalyzer.net/"Mozilla/5.0(Windo  
ws; U; Windows NT 5.2; en-US; rv:1.8.1.7)  
Gecko/20070914 Firefox/2.0.0.7"
```

A user session is a sequence of client request that represent the user's interaction with a web application. Each logged request of the user session is changed into an HTTP request which consists of a URL and name value- pairs to transform log files into test data. Given a set of user sessions, there are techniques of producing test data for web applications [3, 4]. The main points of user session based testing are the selection and reduction methods of test case suites. This paper describes the reduction mechanism of a large number of user session data by applying the clustering algorithm. Representative user sessions are selected from each cluster and then constructed into the test cases for web application. The related work is described in section 2. In section 3, this paper describes dissimilarity function and the clustering algorithm (k-mean-like). Section 4 discusses the proposed system with the metrics for evaluating the cluster