



Relating User Session Clusters to Dynamic Web Application Behavior

Sreedevi Sampath, Sara Sprenkle
University of Delaware

Emily Gibson, Lori Pollock
University of Delaware

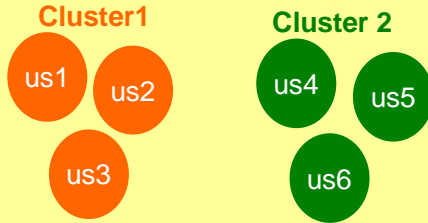
Amie Souter
Drexel University



Goal To study the effectiveness of clustering and test suite reduction, we investigate correlations between user sessions clustered by similar URLs and dynamic program characteristics (program coverage/ fault detection)

Overview

Cluster user sessions, us_i (a sequence of URLs), by similar URLs using a technique called concept analysis



Test suite reduction: our heuristic for reduction selects one user session from each cluster such that sessions in the reduced suite cover all URLs of the application

for example, *Reduced Test Suite: {us2, us6}*

Hypotheses

A cluster of user sessions exhibits similar program characteristics

program code covered/faults detected by a user session, us_i in cluster

measure of similar program characteristics is the common program covered/faults detected by all the sessions in cluster

Cluster 1



Our heuristic selects sessions from clusters that cover different program characteristics

Cluster 1



Cluster 2



program code covered/faults detected by a user session, us_i in cluster

measure of different program characteristics is the difference between common program covered/faults detected within cluster.

Research Questions

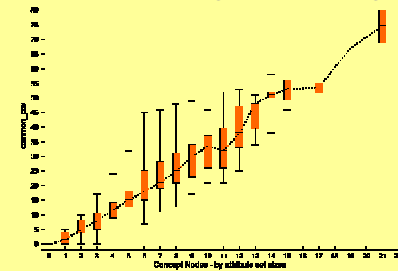
RQ1: How do sessions clustered based on URLs relate to program covered/faults detected by the sessions?

RQ2: Does our heuristic select sessions from clusters that cover different program code and detect different faults?

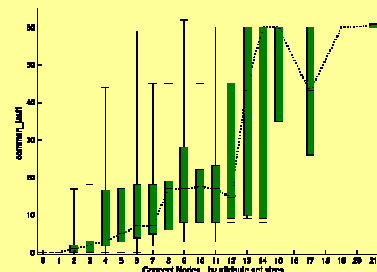
Subject Web Applications and Results

Metrics	Book	CPM
Classes	11	75
Methods	385	172
NCLOC	7791	9298
Faults	40	86
Sessions	125	251

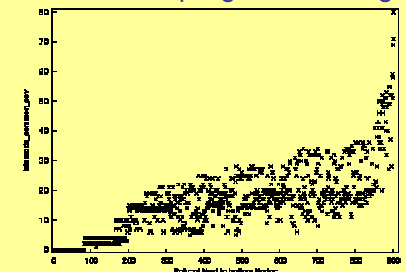
G1: RQ1, program coverage



G2: RQ1, faults detected



G3: RQ2, program coverage



Summary of Results

- As user sessions become more similar in terms of URLs, similarity in their dynamic program characteristics increases (G1, G2)
- Our heuristic selects user sessions from clusters that are different in their program coverage and fault detection (G3)

