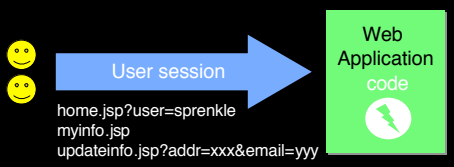
Analyzing Clusters of Web Application User Sessions

Sreedevi Sampath, Sara Sprenkle,
Emily Gibson, Lori Pollock
University of Delaware

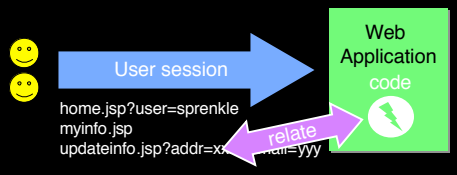
Amie Souter
Drexel University

User-session-based Testing



- User sessions as test cases
  - Real usage of Web applications
  - Beta/maintenance testing phases

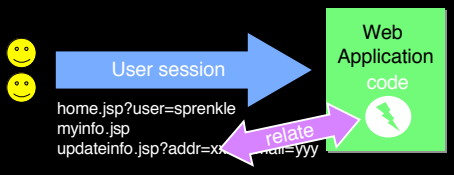
Motivation



Study dynamic behavior of user sessions

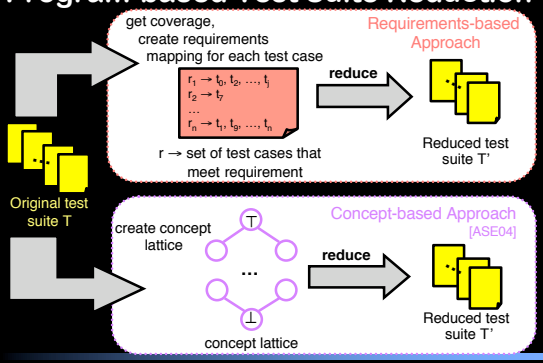
- Correlate user sessions to application behavior
  - program code covered by executing user session
  - faults detected when user session executed

Focus: Test Suite Reduction



- Use correlation to guide testing process
- Key Question: Is there adequate correlation between application behavior and user session clustering characteristics to warrant test suite reduction based on user session attributes only?

Program-based Test Suite Reduction



Concept Analysis

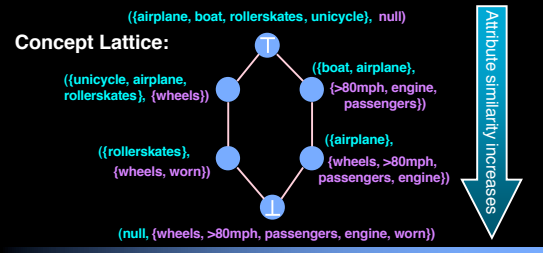
- Mathematical technique for clustering objects by their discrete, common attributes based on a binary relation

Example: Modes of Transportation

objects \ attributes	>80mph	engine	passengers	wheels	worn
Airplane	X	X	X	X	
Boat	X	X	X		
Rollerskates				X	X
Unicycle			X	X	

Concept Analysis

- Mathematical technique for clustering objects by their discrete, common attributes based on a binary relation
- Creates a partial ordering of concepts

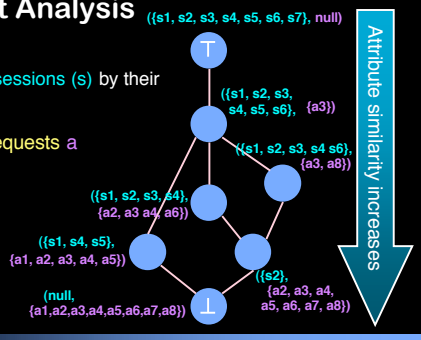


Clustering User Sessions via Concept Analysis

Cluster user sessions (s) by their single URLs

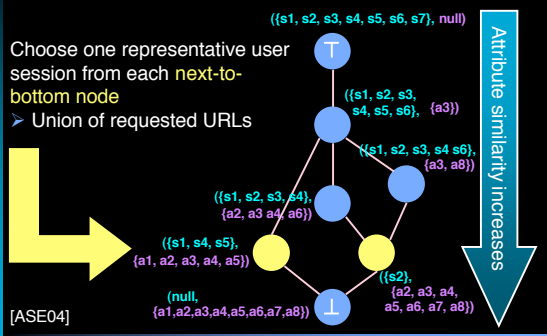
Relation: s requests a

[ASE04]



Test Suite Reduction Heuristic

Choose one representative user session from each next-to-bottom node
- Union of requested URLs



[ASE04]

Clusters and Use Cases

- Characteristics of clusters
  - Shared URLs → use case

User Session B_1
- Search?author=knuth
- ViewItem?id=book023
- X

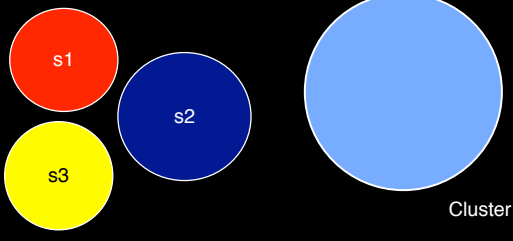
User Session B_2
- Y
- Search?sub=C+programming
- ViewItem?id=book576

Cluster B: Browsing
- Search
- ViewItem

Cluster P: Purchasing
- Search
- ViewItem
- Login
- Purchase

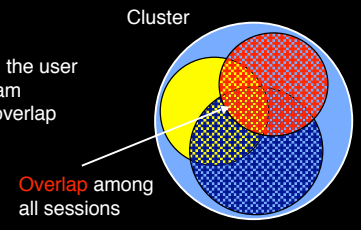
Analyzing Overlap within Clusters

Program characteristics, e.g., program coverage and fault detection, of a cluster's user sessions

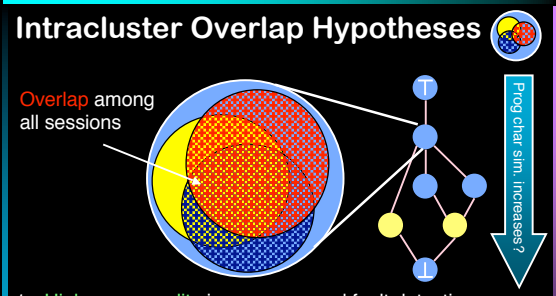


Analyzing Overlap within Clusters

Within a cluster, the user sessions' program characteristics overlap

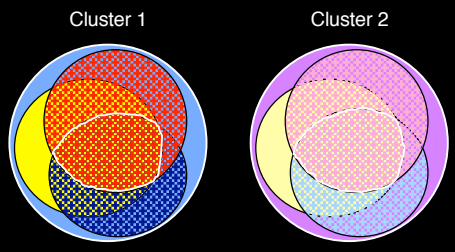


Intracluster Overlap Hypotheses



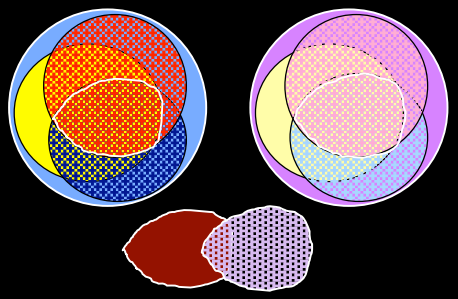
1. High commonality in coverage and fault detection among sessions within a cluster
2. Program-characteristic commonality increases with more common attributes

Analyzing Overlap Across Clusters



Clusters may share common program characteristics

Intercluster Overlap Hypothesis



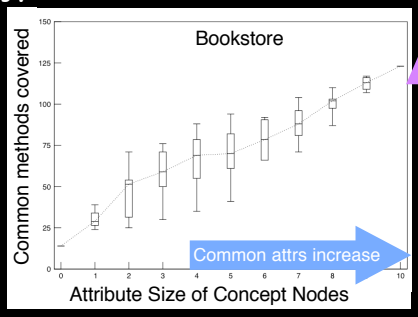
Coverage, detected faults overlap across clusters is small

Case Study Subject Applications

- Bookstore
  - JSPs, MySQL backend
- Course Project Manager (CPM)
  - Java servlets, data store backend

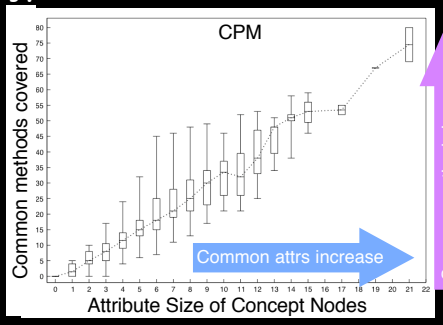
Metrics	Bookstore	CPM
Classes	11	75
Methods	385	172
NCLOC	7791	9298
Faults	40	86
Sessions	125	251

Hypothesis 1



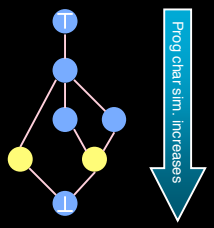
- As attribute set size increases, coverage increases

Hypothesis 1



- Larger lattice, larger attribute sets, similar trend

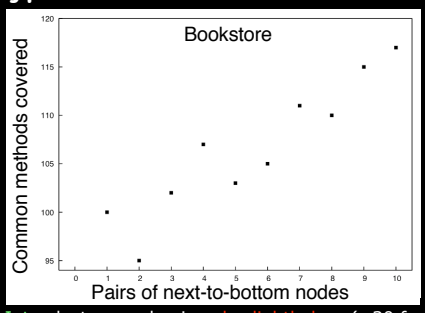
Validation of Intracluster Overlap Hypotheses



- ✓ High commonality in coverage and fault detection among sessions within a cluster
- ✓ Program-characteristic commonality increases with more common attributes

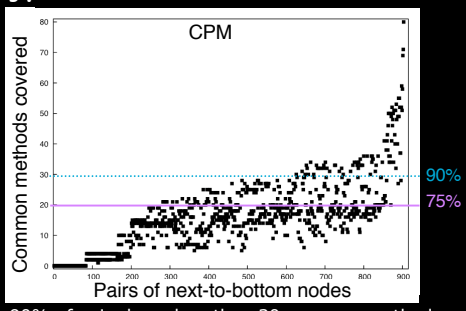
Validates choosing representative user session from cluster

Hypothesis 2



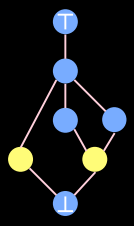
- Intercluster overlap is only slightly less (<30 fewer methods) than intracluster overlap

Hypothesis 2



- 90% of pairs have less than 30 common methods out of 172 total methods

Validation of Intercluster Overlap Hypothesis



✓Program-characteristic commonality small between next-to-bottom nodes for CPM

Validates
- each cluster as a use case
- choosing from next-to-bottom nodes for diverse testing

Summary of Results

- URL similarity implies similarity in program coverage and detected faults
  - Use URLs to describe test case behavior
  - Avoid cost of program-based techniques
- Validates next-to-bottom heuristic
  - A user session is representative of its cluster's program characteristics
  - Next-to-bottom nodes represent different use cases

Future Work

- Additional reduction heuristics
- Alternative clustering techniques

- Discussion: How would clustering by other attributes, e.g., URLs with name-value pairs, change the results for intracluster and intercluster overlap?
  - conclusions about use cases?