Visualization of Program-Execution Data for Deployed Software

Alessandro Orso
James A. Jones
Mary Jean Harrold

Funded by National Science Foundation (CCR-9988294, CCR-0096321, CCR-0205422, SBE-0123532, EIA-0196145) and the State of Georgia through the Yamacraw Mission
Challenges

• Visualization
  • Field data representation
  • Program size
  • Amount of field data

• System for collecting/visualizing data
Outline

• Visualization
  • Field data representation – Color
  • Program size – Visual forms
  • Amount of field data – Visual form, Filters, Summarizers

• System for collecting/visualizing data – Gammatella

• Experience

• Conclusions and Future Work
Color

- Color used to represent characteristics of code and its executions
- Mapping of data to color dependent on intent of visualization

```plaintext
procedure print() {
  for i = 1 to 10 {
    for j = 1 to 10 {
      print j
    }
    print i
  }
}
```
```c
mid() { 
    int x,y,z,m;
    read("Enter 3 numbers:",x,y,z);
    m = z;
    if (y<z) 
        if (x<y) 
            m = y;
        else if (x<z) 
            m = y;
    else 
        if (x>y) 
            m = y;
        else if (x>z) 
            m = x;
    print("Middle number is: ", m);
}
```
Visual Forms – System Level

• For large systems, even File Level view requires scrolling
• Use space-filling representation - TreeMap
Goals

- Preserve relative percentages of colors
- Provide consistent color layout
Goals

- Preserve relative percentages of colors
- Provide consistent color layout
Goals

- Preserve relative percentages of colors
- Provide consistent color layout
Visual Forms - Executions

- Series of vertical bars, each representing an execution

- Large number of executions necessitates richer forms of navigation than simple scroll-bar
Execution Filters

Filter => “User = linus”
Execution Summarizers

- User="linus"
  JVM="jikes1.3"
  Exception="null"

- User="jones"
  JVM="ibm1.4.1"
  Exception="File Not Found"

- User="orso"
  JVM="sun1.4"
  Exception="null"

- User="orso"
  JVM="sun1.4"
  Exception="null"

- User="linus"
  JVM="jikes1.3"
  Exception="Division by Zero"

- User="orso"
  JVM="sun1.4"
  Exception="Array Bounds"

- User="harrold"
  JVM="sun1.3.1"
  Exception="null"

Summarizer => "JVM"

JVM = "jikes1.3"
Gammatella

Software Developer

Program Visualizer

queries

Database

data

InsECT Instrumenter

instrumented program

Data Collection Daemon

execution data

Customer 1

Customer 2

Customer N

At Developers’ Site

In the Field

visualization/interaction

program

queries

execution data
Experience

- Subject program
  - Java Architecture for Bytecode Analysis (JABA)
  - 60,000 lines of code in 550 classes
- Deployment
  - 15 users
  - >2000 executions
- Findings
  - Usage of features
  - Coverage of untested features
  - Usefulness of Gammatella’s features
Gammatella
Related Work

• Eick, Sumner – SeeSoft
• Schneiderman – Treemap
• Baker, Eick – SeeSys

• Reiss, Renieris – Bloom, ALMOST, …
• Jones, Harrold, Stasko – Tarantula
• DePauw et al. – Jinsight
• Jerding, Stasko, Ball – Information Mural
Conclusions

• New approach for visualizing field data
  • Maps field data to program representations using color
  • Provides multiple levels of abstraction for viewing large programs, including new Treemap coloring
  • Accommodates large and continuously increasing amounts of field data using querying techniques
• Toolset, Gammatella, that enables collection and visualization of deployed software’s execution
• Experience with real users and real deployed software
Future Work

• Deploy monitored software to more users to investigate scalability
• Investigate application to exception analysis and fault localization
• Explore other applications of approach such as user-profile extraction
• Investigate interaction and use of public display