Today’s Menu

Last Seminar:
Design & Architecture Reuse
(Tanmay led)

This Seminar:
Reuse Libraries
(Sreekanth is leading)

Next Seminar:
Programming Languages
(Alejandra leads)

Level of Control


Statistical Design

→ Factorial design
  % Two or more factors, e.g., representation method, query set
  % Each factor has discrete values or levels, e.g., (attribute-value, enumerated, faceted, keyword) for "representation method", 4 query sets
  % Factorial points = #levels\(^\text{\text{factor}}\)

→ Learning or order biases
  % Counter: randomize the order in which subjects searched the representation methods

→ Statistical significance
  % if the result is unlikely to have occurred by chance
  % Significance level (critical \(p\)-value): the probability conditional on the null hypothesis of the observed data or more extreme data
  % If the \(p\)-value is small, then it can be said the null hypothesis is false
  % Commonly: \(p<0.05\) (alpha level)
  % ANOVA: analysis of variance

→ Any threats to validity

Psychological Aspect

→ Jon contributed
  % 1234 5678 9012 3456
  % 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
  % Cognition (mental model): combining psychology and SE, Dr. Gary Bradshaw
Library

→ Mitchell Memorial Library
  - A collection of sources, resources, and services, and the structure in which it is housed
  - Organized for use and maintained by a public body, an institution, or a private individual
  - Key stakeholders: creator, user, maintainer...

→ Software Library
  - As old as computing/programming (e.g., COBOL 1959)
  - Static libraries
    - Resolved in a caller at compile-time and copied into a target application by a compiler, producing an object file and a stand-alone executable
    - E.g., `#include <stdio.h>`
  - Dynamic libraries (linking)
    - Loading the subroutines into the application at runtime; the subroutines remain as separate files on disk
    - E.g., DLLs
  - Remote libraries
    - Using completely separate executables and calling them using a RCP (remote procedure call) over a network to another computer
    - E.g., app. server (Enterprise JavaBeans)

Taxonomy


9 Attributes of a software library

- Nature of asset, e.g., source code, req spec., design desc., test data
- Scope of library, e.g., within a project, within an org., world-wide
- Query representation, e.g., keyword, boolean, faceted
- Asset representation, e.g., attribute-value, enumerated, faceted, keyword
- Storage structure, e.g., flat, hypertext, ordering by genericity
- Navigation scheme, e.g., exhaustive linear scan, hypertext links
- Retrieval goal, e.g., correctness (accuracy), functional proximity
- Retrieval criteria, e.g., signature matching, minimizing functional distance
- Matching criteria, e.g., signature identify, NL analysis, pattern recognition

6 software library methods were surveyed and compared

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational Semantics Method</td>
<td>Envisages components</td>
</tr>
<tr>
<td>Information Retrieval Method</td>
<td>Typically within project (application domain)</td>
</tr>
</tbody>
</table>

Summary

→ Reuse Libraries
  - Analogy to (physical) libraries
    - User: browse & retrieval & adapt
    - Maintainer: classify & organize
  - Attributes and methods
  - Advice to practitioners [Frakes-TSE'94]
    - Represent your collection as many ways as you can afford
    - Free text keyword is less expensive (no human indexers needed)
    - Enumerated is the fastest w.r.t. search time
    - None of the 4 methods did better than moderately well

→ Share your domain/topic with us

→ To-do:
  - Project
    - Due in 2 weeks, any progress/problem?
  - Read "Programming Languages" paper