Empirical Study of Representation Methods for Reusable Software Components

William.B.Frakes and Thomas .p.pole
This paper presents

• An empirical study of methods for representing reusable software components.
• Representation methods: keyword, faceted, enumerated and attribute-value
• Tool used: Proteus
Definitions

• Keyword: Term used for words included in a web page that would match words used by web surfers in finding that web page.

• Enumerated: To count off or name one by one; list.

• Attribute value: Define the values of the attribute.

• Faceted: One of numerous aspects. It will allow users ability to find items on more than one dimensions.
Dewey Decimal System

• The Dewey Decimal Classification is a system of library classification made up of ten main classes or categories, each divided into ten secondary classes or subcategories, each having ten subdivisions.
  – 1.1 000 – Computer science, information & general works
  – 1.2 100 – Philosophy and psychology
  – 1.3 200 – Religion
  – 1.4 300 – Social sciences
  – 1.5 400 – Language
  – 1.6 500 – Science
  – 1.7 600 – Technology
  – 1.8 700 – Arts and recreation
  – 1.9 800 – Literature
  – 1.10 900 – History, geography, and biography
Reusable Sw Components

- Reusable software components are designed to apply the power and benefit of reusable, interchangeable parts from other industries to the field of software construction.
- Ex: push buttons, text fields, list boxes, scrollbars, dialogs etc.
Software Reuse

• In most engineering disciplines, systems are designed by composing existing components that have been used in other systems.

• Software engineering has been more focused on original development but it is now recognised that to achieve better software, more quickly and at lower cost, we need to adopt a design process that is based on systematic software reuse
• Part based: The parts-based approach assumes a human programmer integrating software parts into an application by hand.
• Ex: lex and Yacc in Unix Environment.
• Formal Language Based: domain knowledge is encoded into an application generator or a programming language.
• Ex: APL and SAS.
Reuse Benefits

- Increased Dependability
- Reduce Process risk
- Effective use of specialists
- Accelerated Development
Reuse Library

Fig. 1. Reuse library environment.
Reuse Representation Problem

• Methods are drawn from 3 major areas: AI, Hypertext and Library and information science

• Library Issues will be less important in environment with low staff turnover, because component information will be available from people who work in the environment.
Representation methods

Fig. 2. Taxonomy of library science indexing methods.
Controlled vocabulary

- Places limits on the terms that can be used to describe a classified object
- Makes Database Easier to Search
- Ex: Library of Congress Subject, Yellow pages in a phone book
Uncontrolled Vocabulary

• Beetle
• Make-up your face
• Sandwich
• Resume
• SME
Reuse System Studies

• Prieto-Diaz: compared ordered versus unordered keywords; found that the ordered keywords performed better, but no statistical analysis was done.

• Maarek: compared a free text phrase extraction method against free text keyword. They found that the phrases performed better, but provided no statistical analysis to determine if the improvement was significant.
Proteus Experiment

**Proteus Experiment Set Up.**

Instructions on planning an experiment session:

Choose the order in which the methods will be used. As each is chosen, it will be removed from the list of 'Choose the next method:', and added to the list of 'Schedule of Experiment:'.

When all of the methods have been scheduled (chosen), select 'Continue' to start the experiments.

As each method is completed, the user will be returned to this window, where the remaining methods will be displayed in the 'Schedule of Experiment:' window. When 'Continue' is selected again, the next method will be invoked.

<<- | ->>

Schedule of Experiment:

- ENUMERATED-GRAPH
- FACETED-CLASSIFICATION
- KEYWORD-2

CONTINUE

p. 5. Selection screen.
Keyword Search

Fig. 6. Keyword search.
Enumerated Search

Fig. 7. Enumerated search.
Faceted Search

![Faceted Classification Search Window using databases: UNIX-RPG-TOOL](image)

<table>
<thead>
<tr>
<th>PART-NAME</th>
<th>ACTION</th>
<th>FUNCTIONAL-AREA</th>
<th>OBJECT-CLASS-OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>WRITE</td>
<td>ARCHIVAL-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>DR</td>
<td>WRITE</td>
<td>ARCHIVAL-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>ADMIN</td>
<td>DELETE</td>
<td>ARCHIVAL-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>UGCLEANUP</td>
<td>DELETE</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>KEXEC</td>
<td>EXECUTE</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>COMMAND</td>
</tr>
<tr>
<td>RSM</td>
<td>EXECUTE</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>PROCESS</td>
</tr>
<tr>
<td>UUCP</td>
<td>COPY</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>UIDO</td>
<td>SEND</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>FILE</td>
</tr>
<tr>
<td>UXI</td>
<td>EXECUTE</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>COMMAND</td>
</tr>
<tr>
<td>UING</td>
<td>EXECUTE</td>
<td>COMMUNICATIONS-SYSTEM</td>
<td>COMMAND</td>
</tr>
<tr>
<td>STREET</td>
<td>KILL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 8.** Faceted search.
Attribute Value Search

Fig. 9. Attribute value search.
Analysis:

- **Searching Effectiveness**: How well a method supports finding relevant items in a database.
- **Recall**: Number of relevant items retrieved over the number of items in a database.
- **Precision**: Number of items retrieved over the total number of items retrieved.
- **Overlap**: Different methods will retrieve different items.
Fig. 10. Test collection procedure.
Box plot -recall

Fig. 11. Boxplots of recall for the methods.
Fig. 12. Boxplots of precision for the methods.
ANOVA

• Analysis of Variance is a collection of Statistical Models and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables.

• In its simplest form ANOVA gives a statistical test of whether the means of several groups are all equal.
Anova Types

• One way Anova: used to test for differences among two or more independent groups
• Factorial Anova: used when the experimenter wants to study the effects of two or more treatment variables
• Two way Anova: used when the subjects are subjected to repeated measures, in which the same subjects are used for each treatment.
Box plot - Search times

Fig. 13  Boxplots of user search times for the methods.
Conclusion

• There were no significant differences in search effectiveness, as measured by recall and precision, between the four methods.
• Even though the methods were not significantly different in terms of recall and precision, they found different items.
• A library system can be designed to support multiple methods. Proteus shows the technical feasibility of this.
• No method did more than moderately well in terms of search effectiveness as measured by recall and precision.
• Users had no clear preference for a representation method.
Questions

• What is the meaning of * and o notation in figure 13 and figure 15 respectively?
• Why staff turnover period affect the library issue?
• How can this experiment be generalized to industry?
• It seems as if this problem is being approached from a bottom-up manner in which the design of the reuse library and its search methods is defined first and the human subjects are tested against the different designs of the reuse library. Do you think there would be better results if this problem was approached in more of a top-down manner, in which some research is done in the psychology realm to understand how people link search terms together cognitively and then a system was designed to fit the human (instead of the human designed to fit the search methods)?
• One of the authors comments under "Future Work" pertained to replicating this study with "different or larger databases." How could this be done?
Contd....

• What is a repeated measures factorial design?
• Is there any validity threat to this experiment?
• Could you please address a short discussion about ANOVA?
• What does it means that the experiment used a repeated measures factorial design?
• How is the JAD application used?
Why is it important to have a null hypothesis in empirical studies?

What does “Statistically significant” mean?

This paper’s conclusion is largely based on the metrics of recall and precision. However, are they proved to be gold metrics for this experiment?

This paper is more Unix system based, is there any further study on other systems to make the conclusion more generic?
• This paper is an empirical study, which is to say the result inevitably user-biased. Moreover, this paper is more than 10 years ago. Is there any “final” conclusion will the argument in the paper?
• What is scope of this study to replicate with other subjects, different or larger databases, different domains, different interfaces and other representation methods?
On page 6, section C. Statistical Design it says that not all the subjects were given the same set of data to search for. Do you think this takes away from the validity of the experiment since all subjects did not search on the same set of data?
Thank You