N90 - 14790

AUTOMATED REUSABLE COMPONENTS SYSTEM STUDY RESULTS

Kathy Gilroy
Software Productivity Solutions, Inc.

The Automated Reusable Components System (ARCS) was developed under a Phase I Small Business Innovative Research (SBIR) contract for the U.S. Army CECOM. The objectives of the ARCS program were (1) to investigate issues associated with automated reuse of software components, identify alternative approaches, and select promising technologies, and (2) to develop tools that support component classification and retrieval. The approach followed was to research emerging techniques and experimental applications associated with reusable software libraries, to investigate the more mature information retrieval technologies for applicability, and to investigate the applicability of specialized technologies to improve the effectiveness of a reusable component library. Various classification schemes and retrieval techniques were identified and evaluated for potential application in an automated library system for reusable components. Strategies for library organization and management, component submittal and storage, and component search and retrieval were developed. A prototype ARCS was built to demonstrate the feasibility of automating the reuse process. The prototype was created using a subset of the classification and retrieval techniques that were investigated. The demonstration system was exercised and evaluated using reusable Ada components selected from the public domain. A requirements specification for a production-quality ARCS was also developed.

PRECEDING PAGE BLANK NOT FILMED

AUTOMATED REUSABLE COMPONENTS
SYSTEM (ARCS)

Objectives

o Investigate issues associated with automated reuse of software components, identify alternative approaches and select promising technologies
   - classification criteria
   - library organization
   - retrieval techniques

o Develop tools to support component classification and retrieval activities
   - develop demonstration system in Phase I
   - define the requirements for a production-quality system to be developed during Phase II

CLASSIFICATION AND RETRIEVAL RESEARCH
Approach

o Research emerging techniques and experimental applications associated with reusable software libraries

o Investigate the more mature information retrieval (IR) technologies for applicability to the reusable software problem

o Investigate the applicability of specialized technologies (e.g., expert systems, semantic networks, fuzzy logic) to improve the effectiveness of a reusable component library

REUSABLE COMPONENT LIBRARY SYSTEM ROLES

o Classification - the process of entering the component into the library

o Retrieval - the process of finding an applicable component to meet a perceived need

REUSABLE COMPONENT CLASSIFICATION

1. Understand the component

2. Certify the component

3. Classify the component based upon knowledge of the classification strategy

4. Insert the component into the library

REUSABLE COMPONENT RETRIEVAL

1. Access the component by stating the need in terms compatible with the classification system

2. Understand the component

3. Evaluate the component for applicability and acceptability

4. Adapt the component for the particular application

5. Integrate the component into the baseline system under development

Classification Issues

o Classification criteria - those attributes of components that can be used to classify, understand and evaluate them.

o Classification organization - the mechanism by which components are logically organized in the library according to the classification criteria.

Classification Criteria

o Must support both searching and discrimination

o May be static or dynamic

o Classification criteria may be composed of:

- Key words associated with its function, purpose or application area
- Text description
- Characteristics or metrics of interest
- Language or other structured description

Classification Organization

o **Enumerative organizations**

o **Hierarchical taxonomies**

o **Faceted schemes**

o **Semantic nets**

o **Clustered organizations**

Query Logic

o Deterministic logic that retrieves based upon exact matches according to a Boolean query.

o Probabilistic logic that estimates the probability of relevance of specific components in the library.

o Fuzzy logic that uses weighted or graded measures to assess whether a component meets user query criteria.

Query Enhancement

A modification or enhancement of an original query in order to expand or refine the retrieval

o Query Generalization, required when there are too few finds or when the finds that are retrieved are "near- misses".

o Query Specialization, required when user is confronted with too many finds, or when most of the components that are retrieved are non-relevant.

Query Enhancement Application

o Query enhancement experiments in IR have not demonstrated improvements in retrieval effectiveness, and, in fact, demonstrate degradation in many cases

o A compromise to automated query expansion is to calculate an ordering of finds and to use the ordering to present the "best fit" or most relevant to the user first

o Relevance feedback has shown the most promise

Classification and Retrieval Conclusions

No single scheme is best -- employ a number of technologies, adapting and borrowing from database, information retrieval and knowledge-engineering disciplines

o Classification criteria: All types - key words, text, characteristic-based or metrics and languages

o Library classification organization: Faceted, later enhanced with clustering

o Query logic: Deterministic, later enhanced with probabilistic or fuzzy logic

Flexibility is important in improving effectiveness!

ARCS Prototyping Objectives

o Experiment with a faceted classification approach and with supporting multiple classification schemes

o Evaluate candidate criteria for usefulness in retrieval, evaluation and understanding

o Prototype the user interface to improve usability of the production ARCS

o Demonstrate the applicability of an Entity-Relationship database approach

o Support the formulation of requirements for the production ARCS

o Determine areas where more research is needed

Demonstration ARCS Tool

o Implemented entirely in Ada

o Hosted on VAXstation running VMS

o Employed a number of existing components and subsystems

- WINNIE (windowing/menu system)

- SMARSTAR/Rdb (relational database management system)

- Ada Entity-Relationship Interface to database subsystem

- Numerous low level data management components

ARCS Operations

o View the catalog information stored about a specific component existing in the ARCS database.

o Add a catalog entry for a new component, and insert its source code and test cases into the database. This information is then controlled by the ARCS much like checking-in and checking-out information from a CM system.

o Update the catalog information for a specific component.

o Delete all information about an obsolete component from the database.

o Extract a specified component from the database. The sources, tests, and/or catalog information can be copied to a user-specified VMS directory.

o Select (find) components which match search constraints on the values of component characteristics. The components so selected may then be viewed or extracted, or the selection criteria may be modified to improve the results of a subsequent search.

ARCS User Interface

- Consistency, on-line help, shortcuts (data entry still a burden)

- Menus and forms for component attribute update and query

- Supported by windowing and menu organization subsystems

- Attribute-based queries supported by simple query language:

[not]   [qualifier]   value   { $\overset{and}{or}$ [not]   [qualifier] value}

ARCS Database

o Metaschema subschema defines the "super-structure" for the ARCS database, representing the ER model itself.

o Component subschema defines the entities, relationships and attributes containing all of the catalog information about each reusable component.

o Classification subschema defines the entities, relationships and attributes comprising the means for classifying components in different ways.

Implementation Issues Raised and Evaluation Results

o Attribute/criteria selection

o Population of the classification subschema

o Deferred support for certain policies

o Performance Issues

o Data Entry Issues for Usability

14

CONCLUSIONS

o There are sufficient underlying database, IR and knowledge-based technologies on which to develop a production ARCS

o The Phase I research successfully derived a flexible, extensible faceted approach for ARCS and identified promising technologies for further investigation

o The Phase I demonstration system reinforced the validity of the overall approach, while pointing out areas for future investigation

o Additional work is needed to determine the specific classification criteria and classification schemes

o Additional experimentation is needed to address the tradeoffs associated with ease-of-use, performance, applicability and effectiveness