Networks and information technology redefine the practice of reuse

Guillermo Arango

Schlumberger Laboratory for Computer Science
8311 North RR 620, Austin, TX 78720-0015, USA
Tel: 1 (512) 331 3735
Email: arango@austin.slcs.slb.com

Abstract

Libraries, classification schemes, and description languages for software components continue to be central topics of discussion at reuse workshops. I this as a symptom of a community focusing on its own models and beliefs of how the *should* work instead of paying attention to how the world works?

The world uses computer networks for private and public communication. We are now starting to hire from the Nintendo generation. For these young designers and programmers the network *is* the primary source of information. Records of anonymous FTP servers show that the Internet is one of the primary vehicle for finding and distributing software.

The same information services we use to find publications, conference announcements, movie reviews, phone numbers, weather maps, or stock market news are adequate for finding software components. The reuse community should re-examine the relevance of work on libraries and focus on how to leverage off the infrastructural and cultural changes brought about by communications technology.

Keywords: Software component, Component library, Classification, Internet, Information Services, Knowbot, Electronic market.

Workshop Goals: Domain modeling, Reuse of designs, Incremental evolution of components.

1 Background

This position statement is a reaction to what I perceive to be a consistent lack of impact of the output of the reuse research community to the practice of software engineering. Broad-based progress in software reuse appear to be driven not by the results of technical research but by developments in areas such as standards, languages, operating systems, process improvement practices, the Internet, or traditional mail-order operations (*e.g.*, in the US, the Programmer's Connection or Egghead Software catalogues).

2 Position

2.1 The acquisition and reuse of software components is a social process

Work on component reuse assumes that there is a repository and that there is a reuser. That consumers an suppliers do not communicate except through the records in the library and a shared classification scheme. That component classification is needed to reduce search through large collections of components.

In reality,

- Libraries seem to be few and have at most a few hundred components. [Note: we do *not* imply that there are few components available for reuse, there are millions of them available through the (private or public) network infrastructure, but they tend to be distributed through thousands of directories, on thousands of machines.]
- Libraries tend to group components by domains of application and informally shared domain semantics facilitate search through simple catalogue and component inspection.
- In the case of large collections of relatively unstructured information (*e.g.*, FTP directories), there are inexpensive full-text search engines that sift through the material very efficiently. CD Rom technology has lowered the price of software storage and distribution to the point that it is cheap to index every word in a 400 Mb directory and perform almost instant access to any piece of information.
- Information services on the networks multiply. The problem of component classification shifts to *source classification.* Finding useful sources is a more interesting problem than identifying components within a given source.
- Consumer needs are rarely satisfied by looking at existing "libraries". Most people find software through a social process (*e.g.*, bulletin board postings, e-mail) or using information services (*e.g.*, WAIS, Gopher, WWW). Reusers routinely interact with authors via e-mail. Special interest groups share bulletins were reusers post problems and additional advice from members of the group is used by all.

The social process of trading software and software reuse knowledge through the net may not be easy to characterize in formal terms but it is real, it is happening, and it works. Information technology and communication technology are providing the tools to make that process broader and more productive.

2.2 Networks are electronic information markets

Trading software components is just one instance of trading information. If we ignore the information services provided through the network environments we miss a practical, available, and growing infrastructure to enable software access and distribution. This applies to private networks, subscriber networks, and public networks.

At present, a person looking for a component must do most of the foot work: post notes on bulletin boards, browse through FTP directories, Gopher holes. Soon a news group "comp.rfc"-request for component-will facilitate a social search process whereby a posting is responded with direct or indirect references to sources of the component. Requests and components will come from individuals in different organizations in different countries. Should we focus on the problem of matching requests and offers that use an un-constrained vocabulary (assuming most people will use English)? or should we promote standard templates?

In the electronic information markets that are emerging today (*e.g.*, 170.000 postings a day on Prodigy alone), we anticipate a more balanced role between consumers and suppliers. Suppliers for profit will also search for requests-for-components that match their inventory or their capabilities to supply the components through programming, customization, or generation. The distinction between reuse and outsourcing will blur. From the point of view of the consumer the genesis of the component will not matter if the economics, documentation, and QA are right. Suppliers of custom components will themselves outsource to suppliers in regions with inexpensive labor costs. Some consumers will engage in supplier qualification and pre-qualification. [By the way,... this is already happening today both for software and for hardware components. With networks as primary communication media, and EDI and document format interchange standards (e.g., JPEG, MPEG, SGML, HTML) as enablers.]

Locating "useful" components for specific projects and product lines will not be a one-shot effort, it will become part of a normal *technology watch* activity, with knowbots [R. Kahn, Center for National Research Initiatives, 1988] and information filters screening information through bulletin-boards, news bulletins, and negotiating access to information with information-brokerage applications. [Again... for non-software types of information, these capabilities are commercially available to subscribers today.]

With time we expect that the search for customers will also become an active, permanent activity Knowbots with access to component descriptors will also monitor network traffic to identify potential consumers. In effect, we will have components looking for reusers.

2.3 Reuse processes and problems are being reformulated

The reuse research community should pay attention to what information technology and networks are offering to the common person. The evolution of these technologies is altering the way people communicate and work. As a consequence, many of the assumptions about the classic "reuse process" are not valid any longer and new opportunities are emerging. For example, the assumption about the library and the classification scheme being the only communication media between suppliers and consumers is not realistic; reuse processes based on N^2 communication between authors and reusers of a component take place routinely via e-mail and must be studied..

Some old problems, such as component classification or library search are being reformulated or disposed. For example, the ROI of existing and new approaches to component classification, schema evolution, and library search should be benchmarked against the ROI of full-text search on normal directories. Only those that can outperform brute-force search substantially should be retained. Another example is the trade-off between cost (and actual effectiveness) of classification schemes to enable automated searches to "hit" on the right component description vs. brute-force search coupled with a good standard for component documentation.

The wealth of sources and resources available on the networks today is such that there are many new problems, for example, How to develop applications that monitor known network sources and find new sources, How to program an application to monitor network traffic for specific types of information and draw useful inferences from pieces obtained from heterogeneous sources, How to structure repository "yellow pages" in the Internet, How to standardize on component documentation and characterization for an open market.

And there is also an opportunity to focus on some of the *essential* software engineering aspects of reuse, for example, How to facilitate understanding a component, How to define its *design envelope*, How to identify and propagate changes reliably as a component is ported, customized, integrated into a system, or optimized.

3 Biography

Guillermo Arango is Research Scientist and Manager of the Design Program at the Schlumberger Laboratory for Computer Science in Austin,TX. The Program conducts research on design process, methods and tools in software and mechanical engineering, information management in the design process, electronic publishing of technical documentation, and applications of artificial neural networks to data interpretation.

Guillermo has worked in software engineering for the past 15 years and on reuse-related topics for the past 10 years. Five of them in a university environment (code generation, automated software evolution, domain modeling) and five of them in industry. His interests include: domain analysis, design analysis, reuse of designs, management of change, design rationales, management and visualization of design information. His work applies to product families involving multiple technologies (*e.g.*, software, mechanics, electronics, hydraulics).