A Procedure for Evaluating Human-Computer Interface Development Tools

By Deborah Hix

TR 89-29

Proceedings of the Second UIST Conference,
November 1989, Williamsburg, VA.

A PROCEDURE FOR EVALUATING HUMAN-COMPUTER INTERFACE DEVELOPMENT TOOLS

Deborah Hix
Department of Computer Science
Virginia Polytechnic Institute & State University
Blacksburg, VA 24061
703/231-6199
email: hix@vtodie.cs.vt.edu

1. INTRODUCTION

1.1. Motivation

The computer industry has seen an explosive emergence of human-computer interface development tools -- sometimes called user interface management systems or UIMS -- in the last few years. Commercial software packages and research systems even tangentially related to the area of human-computer interaction now claim to be UIMS. Human-computer interface development tools are themselves interactive systems that support production and execution of the human-computer interface. With their recent proliferation, evaluations and comparisons are constantly done, but without a formal, structured approach. State-of-the-art in evaluation of these tools is based mostly on warm fuzzy feelings and the ever-popular opinion. Addressing these problems is difficult, largely because of the relative newness of such tools, because of the many differents kinds of systems that are called UIMS, and because of their inherent complexity. These tools are complex because human-computer interfaces, which these tools produce, are complex.

An evaluation procedure that uses a standardized technique to produce quantifiable criteria for evaluating and comparing human-computer interface development tools is described in this paper. An empirical validation study, to determine the consistency of ratings produced by this procedure, is also presented. These ratings could be used, for example, as important data for the task of choosing a tool for a particular human-computer interface development environment. The research presented in this paper is one of the first attempts to produce a structured, quantitative approach to evaluating such tools. Our procedure is being used in several commercial interface development environments (e.g., GTE Data Systems [Arble 1988] and McDonnell Douglas [Totten 1989]).

1.2. Related Research

Precedence has been established for using the type of approach on which our tool evaluation procedure is based. Roberts and Moran [1983], for example, produced a methodology for standardized evaluation of text editors. The basis of their approach was classification of potential editing tasks and evaluation along several dimensions, including time to perform tasks, error costs, learning time, and functionality. A replication study of their work [Borenstein 1985] produced recommendations for modifications, including limitations on using a stopwatch, changes in testing expertise, and extension of the functionality dimension. Cohill, Gilfoil, and Pilitsis [1988] developed a methodology used at AT&T for evaluating software packages, particularly commercial systems. Criteria such as performance, documentation, and support were evaluated. Research such as this provided ideas used in the development of our tool evaluation procedure.