

Abstract

An Empirical Foundation for Automated
Web Interface Evaluation

by

Melody Yvette Ivory
Doctor of Philosophy in Computer Science

University of California at Berkeley

Professor Marti Hearst, Chair

This dissertation explores the development of an automated Web evaluation methodology and tools. It presents an extensive survey of usability evaluation methods for Web and graphical interfaces and shows that automated evaluation is greatly underexplored, especially in the Web domain.

This dissertation presents a new methodology for HCI: a synthesis of usability and performance evaluation techniques, which together build an empirical foundation for automated interface evaluation. The general approach involves: 1. identifying an exhaustive set of quantitative interface measures; 2. computing measures for a large sample of rated interfaces; 3. deriving statistical models from the measures and ratings; 4. using the models to predict ratings for new interfaces; and 5. validating model predictions.

This dissertation presents a specific instantiation for evaluating information-centric Web sites. The methodology entails computing 157 highly-accurate, quantitative page-level and site-level measures. The measures assess many aspects of Web interfaces, including the amount of text on a page, color usage, and consistency. These measures along with expert ratings from Internet professionals are used to derive statistical models of highly-rated Web interfaces. The models are then used in the automated analysis of Web interfaces.

This dissertation presents analysis of quantitative measures for over 5300 Web pages and 330 sites. It describes several statistical models for distinguishing good, average, and poor pages with 93%–96% accuracy and for distinguishing sites with 68%–88% accuracy.

This dissertation describes two studies conducted to provide insight about what the statistical models assess and whether they help to improve Web design. The first study attempts to link expert ratings to usability ratings, but the results do not enable strong conclusions to be drawn. The second study uses the results of applying the statistical models for assessing and refining example sites and shows that pages and sites modified based on the models are preferred by participants – professional and non Web designers – over the original ones. Finally, this dissertation demonstrates use of the statistical models for assessing existing Web design guidelines.

This dissertation represents an important first step towards enabling non-professional designers to iteratively improve the quality of their designs.