

# An Empirical Foundation for Automated Web Interface Evaluation

by

Melody Yvette Ivory

B.S. (Purdue University) 1993

M.S. (University of California at Berkeley) 1996

A dissertation submitted in partial satisfaction of the  
requirements for the degree of  
Doctor of Philosophy

in

Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Marti Hearst, Chair

Professor James Landay

Professor Ray Larson

Fall 2001

**An Empirical Foundation for Automated  
Web Interface Evaluation**

Copyright Fall 2001  
by  
Melody Yvette Ivory

## 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.

To Alluminum and Professo.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Usability Evaluation of User Interfaces</b>	<b>4</b>
2.1	Introduction . . . . .	4
2.2	The Usability Evaluation Process . . . . .	6
2.2.1	Specify Usability Evaluation Goals . . . . .	6
2.2.2	Determine UI Aspects to Evaluate . . . . .	7
2.2.3	Identify Target Users . . . . .	7
2.2.4	Select Usability Metrics . . . . .	7
2.2.5	Select Evaluation Method(s) . . . . .	7
2.2.6	Select Tasks . . . . .	8
2.2.7	Design Experiments . . . . .	8
2.2.8	Capture Usability Data . . . . .	8
2.2.9	Analyze and Interpret Data . . . . .	9
2.2.10	Critique UI to Suggest Improvements . . . . .	9
2.2.11	Iterate Process . . . . .	9
2.2.12	Present Results . . . . .	9
2.3	Taxonomy of Usability Evaluation Automation . . . . .	9
2.3.1	Proposed Taxonomy . . . . .	10
2.4	Overview of Usability Evaluation Methods . . . . .	12
2.5	Usability Testing Methods . . . . .	15
2.5.1	Usability Testing Methods: Non-automated . . . . .	15
2.5.2	Usability Testing Methods: Automated Capture . . . . .	18
2.5.3	Usability Testing Methods: Automated Analysis . . . . .	22
2.6	Inspection Methods . . . . .	30
2.6.1	Inspection Methods: Non-automated . . . . .	30
2.6.2	Inspection Methods: Automated Capture . . . . .	32
2.6.3	Inspection Methods: Automated Analysis . . . . .	32
2.6.4	Inspection Methods: Automated Critique . . . . .	35
2.7	Inquiry Methods . . . . .	37
2.7.1	Inquiry Methods: Non-automated . . . . .	37
2.7.2	Inquiry Methods: Automated Capture . . . . .	39
2.8	Analytical Modeling Methods . . . . .	39
2.8.1	Analytical Modeling Methods: Non-automated . . . . .	40
2.8.2	Analytical Modeling Methods: Automated Analysis . . . . .	41
2.9	Simulation Methods . . . . .	43
2.9.1	Simulation Methods: Automated Capture . . . . .	44

2.9.2	Simulation Methods: Automated Analysis . . . . .	45
2.10	Expanding Existing Approaches to Automating Usability Evaluation Methods . . . .	48
2.10.1	Expanding Log File Analysis Approaches . . . . .	49
2.10.2	Expanding Guideline Review Approaches . . . . .	50
2.10.3	Expanding Analytical Modeling Approaches . . . . .	51
2.10.4	Expanding Simulation Approaches . . . . .	51
2.11	Summary . . . . .	53
<b>3</b>	<b>New Automated Usability Evaluation Methods</b>	<b>54</b>
3.1	Introduction . . . . .	54
3.2	The Performance Evaluation Process . . . . .	54
3.3	Overview of Performance Evaluation Methods . . . . .	55
3.4	Measurement Methods . . . . .	56
3.4.1	Measurement Methods: Selecting Performance Metrics . . . . .	56
3.4.2	Measurement Methods: Running a Workload . . . . .	57
3.4.3	Measurement Methods: Capturing Performance Metrics . . . . .	59
3.5	Analytical Modeling Methods . . . . .	59
3.6	Simulation Methods . . . . .	60
3.7	Mapping Between Performance and Usability Evaluation . . . . .	61
3.7.1	Example Applications and Tasks . . . . .	62
3.8	New UE Measurement Methods . . . . .	63
3.8.1	Measurement in Performance Evaluation . . . . .	63
3.8.2	Measurement in Usability Evaluation . . . . .	64
3.8.3	Applying PE Measurement Methods to UE . . . . .	65
3.8.4	Example Usability Benchmark . . . . .	66
3.9	New UE Analytical Modeling Methods . . . . .	69
3.9.1	Analytical Modeling in Performance Evaluation . . . . .	69
3.9.2	Analytical Modeling in Usability Evaluation . . . . .	69
3.9.3	Applying PE Analytical Modeling to UE . . . . .	69
3.9.4	Example Analytical Models . . . . .	71
3.10	New UE Simulation Methods . . . . .	72
3.10.1	Simulation in Performance Evaluation . . . . .	72
3.10.2	Simulation in Usability Evaluation . . . . .	72
3.10.3	Applying PE Simulation to UE . . . . .	73
3.10.4	Example Simulators . . . . .	74
3.11	Summary . . . . .	77
<b>4</b>	<b>Automated Analysis Methodology and Tools</b>	<b>79</b>
4.1	Introduction . . . . .	79
4.2	Analysis Scenarios . . . . .	79
4.2.1	Web Interface Evaluation . . . . .	79
4.2.2	Web Interface Design . . . . .	80
4.3	Methodology . . . . .	81
4.4	Tools . . . . .	83
4.4.1	HTML Parser and Browser Emulator . . . . .	84
4.4.2	Site Crawler Tool . . . . .	85
4.4.3	Metrics Computation Tool . . . . .	85
4.4.4	Analysis Tool . . . . .	86

4.5	Summary . . . . .	86
<b>5</b>	<b>Web Interface Measures</b>	<b>88</b>
5.1	Introduction . . . . .	88
5.2	Web Interface Structure . . . . .	88
5.3	Summary of Web Interface Measures . . . . .	90
5.4	Text Element Measures . . . . .	92
5.4.1	Text Element Measures: Page Text . . . . .	92
5.4.2	Text Element Measures: Page Title . . . . .	96
5.4.3	Text Element Measures: Page Abstract . . . . .	96
5.4.4	Text Element Measures: Body Text . . . . .	96
5.4.5	Text Element Measures: Link Text . . . . .	97
5.4.6	Text Element Measures: Link Text Length . . . . .	97
5.4.7	Text Element Measures: Content Percentage . . . . .	98
5.4.8	Text Element Measures: Navigation Percentage . . . . .	98
5.4.9	Text Element Measures: Exclamation Points . . . . .	98
5.4.10	Text Element Measures: Typographical Errors . . . . .	98
5.4.11	Text Element Measures: Readability . . . . .	99
5.4.12	Text Element Measures: Information Quality . . . . .	100
5.5	Link Element Measures . . . . .	100
5.5.1	Link Element Measures: Links . . . . .	100
5.5.2	Link Element Measures: Text Links . . . . .	102
5.5.3	Link Element Measures: Link Graphics . . . . .	102
5.5.4	Link Element Measures: Within-Page Links . . . . .	102
5.5.5	Link Element Measures: External Links . . . . .	102
5.5.6	Link Element Measures: Embedded Links . . . . .	103
5.5.7	Link Element Measures: Wrapped Links . . . . .	103
5.5.8	Link Element Measures: Redundant Links . . . . .	103
5.5.9	Link Element Measures: Navigation Quality . . . . .	103
5.6	Graphic Element Measures . . . . .	104
5.6.1	Graphic Element Measures: Graphics . . . . .	104
5.6.2	Graphic Element Measures: Graphical Links . . . . .	106
5.6.3	Graphic Element Measures: Graphical Ads . . . . .	106
5.6.4	Graphic Element Measures: Animation . . . . .	107
5.6.5	Graphic Formatting Measures: Graphic Quality . . . . .	107
5.7	Text Formatting Measures . . . . .	107
5.7.1	Text Formatting Measures: Text Emphasis . . . . .	108
5.7.2	Text Formatting Measures: Body Text Emphasis . . . . .	111
5.7.3	Text Formatting Measures: Font Styles . . . . .	111
5.7.4	Text Formatting Measures: Font Point Sizes . . . . .	112
5.7.5	Text Formatting Measures: Text Colors . . . . .	112
5.7.6	Text Formatting Measures: Text Positioning . . . . .	113
5.7.7	Text Formatting Measures: Text Clustering . . . . .	113
5.7.8	Text Formatting Measures: Lists . . . . .	114
5.7.9	Text Formatting Measures: Rules . . . . .	114
5.7.10	Text Formatting Measures: Text in Clusters . . . . .	114
5.8	Link Formatting Measures . . . . .	114
5.8.1	Link Formatting Measures: Non-Underlined Links . . . . .	115

5.8.2	Link Formatting Measures: Link Colors . . . . .	115
5.9	Graphic Formatting Measures . . . . .	115
5.10	Page Formatting Measures . . . . .	116
5.10.1	Page Formatting Measures: Colors . . . . .	117
5.10.2	Page Formatting Measures: Color Combinations . . . . .	121
5.10.3	Page Formatting Measures: Fonts . . . . .	122
5.10.4	Page Formatting Measures: Line Length . . . . .	122
5.10.5	Page Formatting Measures: Leading . . . . .	122
5.10.6	Page Formatting Measures: Framesets . . . . .	122
5.10.7	Page Formatting Measures: Interactive Elements . . . . .	123
5.10.8	Page Formatting Measures: Screen Size . . . . .	123
5.10.9	Page Formatting Measures: Screen Coverage . . . . .	124
5.10.10	Page Formatting Measures: Text Density . . . . .	124
5.10.11	Page Formatting Measures: Scrolling . . . . .	124
5.10.12	Page Formatting Measures: Stylesheets . . . . .	124
5.10.13	Page Formatting Measures: Layout Quality . . . . .	125
5.11	Page Function . . . . .	125
5.12	Page Performance Measures . . . . .	126
5.12.1	Page Performance Measures: Page Bytes . . . . .	128
5.12.2	Page Performance Measures: Graphic Bytes . . . . .	133
5.12.3	Page Performance Measures: Objects . . . . .	133
5.12.4	Page Performance Measures: Download Speed . . . . .	134
5.12.5	Page Performance Measures: Accessibility . . . . .	135
5.12.6	Page Performance Measures: HTML Errors . . . . .	135
5.12.7	Page Performance Measures: Scent Quality . . . . .	136
5.13	Site Architecture Measures . . . . .	138
5.13.1	Site Architecture Measures: Consistency . . . . .	138
5.13.2	Site Architecture Measures: Size . . . . .	141
5.14	Summary . . . . .	142
<b>6</b>	<b>Profiles of Highly-Rated Web Interfaces</b>	<b>143</b>
6.1	Introduction . . . . .	143
6.2	Background: Prior Profile Development Work . . . . .	143
6.3	Data Collection . . . . .	144
6.3.1	The Webby Awards 2000 . . . . .	144
6.3.2	Analysis Data . . . . .	145
6.4	Profile Development Methodology . . . . .	146
6.5	Summary of Developed Profiles . . . . .	147
6.5.1	Page-Level Models . . . . .	147
6.5.2	Page-Level Models: Key Predictor Measures . . . . .	149
6.5.3	Site-Level Models . . . . .	154
6.5.4	Site-Level Models: Key Predictor Measures . . . . .	154
6.6	Overall Page Quality . . . . .	155
6.6.1	Overall Page Quality: Classification Model . . . . .	155
6.6.2	Overall Page Quality: Characteristics of Good, Average, and Poor Pages . . . . .	156
6.7	Good Page Clusters . . . . .	161
6.8	Page Type Quality . . . . .	165
6.9	Content Category Quality (Pages) . . . . .	173

6.10	Overall Site Quality . . . . .	175
6.11	Content Category Quality (Sites) . . . . .	178
6.12	Summary . . . . .	178
<b>7</b>	<b>Linking Web Interface Profiles to Usability</b>	<b>180</b>
7.1	Introduction . . . . .	180
7.2	User Study Design . . . . .	181
7.2.1	Study Sites and Tasks . . . . .	181
7.2.2	Participants . . . . .	185
7.2.3	Testing Session . . . . .	186
7.2.4	Testing Environment . . . . .	188
7.3	Data Collection . . . . .	188
7.3.1	Basic Information . . . . .	191
7.3.2	Objective Measures . . . . .	191
7.3.3	Subjective Measures . . . . .	191
7.3.4	Screening of Objective and Subjective Measures . . . . .	192
7.4	Developing a Composite WAMMI Score . . . . .	192
7.5	Mapping Between WAMMI Scales and Webby Scores . . . . .	193
7.6	Perceived Usability Results . . . . .	196
7.6.1	Perceived Usability of Good and Poor Sites (Objective and Subjective Measures) . . . . .	196
7.6.2	Consistency of Perceived Usability Ratings and Webby Scores (Subjective Measures) . . . . .	197
7.7	Actual Usability Results . . . . .	197
7.7.1	Actual Usability of Good and Poor Sites (Objective and Subjective Measures) . . . . .	198
7.7.2	Consistency of Actual Usability Ratings and Webby Scores (Subjective Measures) . . . . .	199
7.8	Summary . . . . .	199
<b>8</b>	<b>Applying the Web Interface Profiles: Example Web Site Assessment</b>	<b>200</b>
8.1	Introduction . . . . .	200
8.2	The Example Site . . . . .	200
8.3	Site-Level Assessment . . . . .	204
8.4	Page-Level Assessment . . . . .	204
8.5	Summary of Assessment Findings . . . . .	209
8.6	Improving the Site . . . . .	213
8.7	Summary . . . . .	218
<b>9</b>	<b>Evaluating the Web Interface Profiles</b>	<b>219</b>
9.1	Introduction . . . . .	219
9.2	Study Design . . . . .	219
9.2.1	Study Sites . . . . .	219
9.2.2	Participants . . . . .	226
9.2.3	Testing Interface . . . . .	234
9.2.4	Testing Session . . . . .	234
9.2.5	Testing Environment . . . . .	236
9.3	Data Collection . . . . .	238
9.4	Page-Level Results . . . . .	238

9.5	Site-Level Results . . . . .	239
9.6	Summary . . . . .	241
<b>10</b>	<b>Applying the Web Interface Profiles: Empirical Examination of Web Design Guidelines</b>	<b>243</b>
10.1	Introduction . . . . .	243
10.2	Page-Level Guidelines . . . . .	244
10.2.1	Amount of Text on a Page (Text Element) . . . . .	244
10.2.2	Length and Quality of Link Text (Text Element) . . . . .	245
10.2.3	Number and Type of Links on a Page (Link Element) . . . . .	245
10.2.4	Use of Non-Animated and Animated Graphical Ads (Graphic Element) . . . . .	246
10.2.5	Font Styles and Sizes (Text and Page Formatting) . . . . .	247
10.2.6	Unique Colors and Color Combinations (Text, Link, and Page Formatting) . . . . .	248
10.2.7	Download Speed (Page Performance) . . . . .	249
10.2.8	Accessibility and HTML Errors (Page Performance) . . . . .	250
10.3	Site-Level Guidelines . . . . .	251
10.3.1	Consistency Across Pages (Site Architecture) . . . . .	251
10.4	Summary . . . . .	251
<b>11</b>	<b>Conclusions and Future Work</b>	<b>253</b>
<b>A</b>	<b>Automation Characteristics of UE Methods, Chapter 2</b>	<b>273</b>
A.1	Automation Characteristics of WIMP UE Methods . . . . .	273
A.2	Automation Characteristics of Web UE Methods . . . . .	275
<b>B</b>	<b>Label Creation Task, Chapter 3</b>	<b>279</b>
B.1	Complete Task Sequence . . . . .	279
B.2	High-Level Task Sequence . . . . .	279
<b>C</b>	<b>Benchmarking Tools and Web Interface Measures, Chapters 4 and 5</b>	<b>284</b>
C.1	Benchmarking Tools . . . . .	284
C.2	Web Interface Measures . . . . .	284
<b>D</b>	<b>Comparison of Original and Modified Web Pages, Chapters 8 and 9</b>	<b>285</b>
D.1	Web Pages from the Example Assessment . . . . .	285
D.2	Web Pages from the Profile Evaluation Study . . . . .	285

## Acknowledgements

First and foremost, all thanks to the most high for giving me the strength, determination, and courage to weather the many storms encountered along the way to completing this dissertation.

I want to thank Alluminum (Allum Ross) and Professo (Angelo Ivory) for putting up with me even at my worst. Your patience, love, and support day in and day out brought me through. Whatever I needed you to do and be, you were there for me. Thank you for taking better care of me than I could possibly take of myself. This dissertation would not be possible without you, so I dedicate it to you. I love you from the bottom of my heart!

I want to acknowledge my advisor, Marti Hearst, for allowing me the freedom to walk this path in my own way while simultaneously shaping and molding me to tap into my full potential. You have been a friend, a mentor, a mother at times, and so much more. You are a tough cookie, but I am glad that I had the privilege of working with you. I want you to know that I truly appreciate you (even your unreasonable demands and high expectations)!

I also want to thank my first advisor, Jim Demmel, for getting me started on this path. I appreciate you taking the risk to have me as your student and teaching me about research. What I learned from our work together played a major role in this dissertation.

I want to acknowledge my other dissertation committee members, James Landay and Ray Larson, for reading this tome in record time so I could bring this journey to an end. I greatly appreciate your support!

I want to thank Professor Kahan for taking interest in my success as a graduate student as well as my effectiveness as a parent. I greatly appreciate your input and guidance over the years.

I want to acknowledge Dr. Sheila Humphreys for being my miracle worker over the years. No matter what crisis came up, you waved your magic wand to make it go away. Even at the last minute when I spilled juice on my laptop and rendered it useless, you loaned me your personal laptop so I could continue pushing forward. I cannot even begin to express my gratitude for your being the wonderful person that you are. How can I ever repay you?

I want to thank Dr. Rashmi Sinha for helping me to lay the foundation for this dissertation. You taught me a lot of what I know about advanced statistics, and this dissertation would not have been possible without your input. Thank you for allowing me to call you with hard questions whenever I needed to. Furthermore, thank you for being a friend even when I was not.

I want to acknowledge all of the GUIR people for your invaluable input over the years. I want to acknowledge the undergraduate and graduate researchers that have contributed to my research over the years: Deep Debroy, Stephen Demby, David Lai, Wai-ling Ho-Ching, Toni Wadjiji, and several others. I want to acknowledge Lincoln Stein for his assistance with an early version of the Metrics Computation Tool, Tom Phelps for his assistance with the extended Metrics Computation Tool, and Maya Draisin and Tiffany Shlain at the International Academy of Digital Arts and Sciences for making the Webby Awards 2000 data available. I want to acknowledge Jared Spool and User Interface Engineering for providing a complimentary copy of the *Designing Information-Rich Web Sites* report series and Doug van Duyne and others at NetRaker for making it possible to use their Web research tools. Finally, I want to acknowledge other researchers in the HCI community who have provided valuable input into this research, including Ben Shneiderman, Jean Scholtz, Sharon Laskowski, B.J. Fogg, and Jean Vanderdonckt.

I want to thank everyone within Kaiser Permanente and especially within the Web Portal Services Group for your unfaltering support of my completing this dissertation. I want to thank my manager, Mary-Anna Rae, for being so understanding, patient, and supportive. I want to thank everyone for graciously giving up their time to participate in the final user study, including Dani Tobler, Mark Randall, Arturo Delgado, Kayvan Sootodeh, Royce Everone, Gaston Cangiano,

Bryan Merrill, Lisa Washington, Katrina Rosario, Anthony Harris, Mary-Anna Rae, Santhosh Pillai, Cheryl Rucks, Kelly Albin, and Barbara Andrews. I want to thank Ron Dolin for bringing me into this special group, and the upper management team for making it possible for me to remain a part of this group – Doug Shelton, Ric Leopold, Kayvan Sotoodeh, Ben Say, and Dr. Henry Neidermeier. I also want to thank the many clients that I have worked with; these assignments helped me to hone my skills as a usability engineer and to put my money where my mouth is.

I want to thank Reggie Walker for helping me to get into graduate school in the first place. I want to acknowledge the family members and friends who played a role in my successfully completing this program: my brother LaVance and his family, my niece Alisia Ivory, Ronald Ringo and family, Damesha Ringo, Bertrand Gaskin, Teddy and Teresa Gaskin, Lin Covington, Janice Sherman, Carmen Anthony, Richard Freeman, Dr. Carla Trujillo, Dr. Johnathan Reason, Dr. Adrian Isles, Dr. John Davis, Roy Sutton, and other members of BGESS. I want to acknowledge the faculty and staff at the Head-Royce school for making it possible for my son to receive a quality education. If I have left anyone out, I do apologize. Please know that you are certainly not forgotten. I am just all thought out right now.

Finally, yet importantly, I want to acknowledge the funding sources that made it possible to complete this program: a Microsoft Research Grant, a Gates Millennium Fellowship, a GAANN fellowship, a Lucent Cooperative Research Fellowship Program grant, an AT&T Cooperative Research Fellowship Program grant, an NSF Honorable Mention grant, and others. I want to thank John Choppy in the Berkeley financial aid office for helping me to navigate the financial aid muddle over the years. I want to acknowledge Professor Charles Thompson from the AT&T and Lucent Cooperative Research Fellowship programs for challenging me to finish my program. When can I expect my Montblanc pen?