

UNDERSTANDING IMPLICIT FEEDBACK AND DOCUMENT PREFERENCE:

A NATURALISTIC USER STUDY

BY

JOYCE DIANE KELLY

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Communication, Information and Library Studies

written under the direction of

Nicholas J. Belkin

and approved by

New Brunswick, New Jersey

January, 2004

ABSTRACT OF THE DISSERTATION

Understanding Implicit Feedback and Document Preference: A Naturalistic User Study

by JOYCE DIANE KELLY

Dissertation Director:

Nicholas J. Belkin

As the amount of online information increases every day, tailoring system responses to individual interests is becoming an important problem in information systems research. This dissertation seeks to understand how an online information system can automatically predict which Web documents users prefer by monitoring their online behaviors with documents. The dissertation is further concerned with understanding how these behaviors are related to the context in which users seek information. Behaviors under investigation include how long users display documents in their browser windows (display time) and if users save, print or bookmark documents (retention). A user's preference for a document is measured by a user-assigned usefulness score. Information-seeking context is characterized by users' self-identified tasks and topics, and several attributes of these, such as the length of time the user expects to be working on a task and the user's familiarity with a topic. To observe users in natural information-seeking situations, users were provided with new laptops and printers, and their online interactions were unobtrusively monitored for fourteen weeks with client- and proxy-side logging software. At weekly intervals, subjects evaluated the usefulness of the documents that they viewed, classified these documents according to their tasks and

topics, and characterized other information-seeking context attributes. Results demonstrated no direct relationship between display time and usefulness, and that display time was significantly related to information-seeking context in different ways, for different subjects. Most notably, display times differed significantly according to task and topic, and topic familiarity. In addition, retention was not always a good indicator of document preference, and subjects were more likely to retain documents related to particular tasks and topics. Finally, results showed no correlation between proxy- and client-generated display times, and that erroneous results are likely to occur when using proxy-generated display times. Overall, these results indicate that for an online information system to use behaviors to infer document preference, it is necessary for the system to model the user's information-seeking context, and that approaches to modeling should be personal rather than general. Furthermore, the integrity of behavior-based metrics used by such systems is an important issue that deserves special attention.

ACKNOWLEDGEMENTS

I would most like to thank Nick Belkin, my mentor and advisor, for his continued support, encouragement and guidance, both intellectually and socially, for which I am indebted forever. I would also like to thank the members of my dissertation committee, Tefko Saracevic, Anselm Spoerri, Doug Riecken and Gene Golovchinsky, for their contributions and support. I am especially grateful to Gene, who gave me the opportunity to work with him one very short, but fun, summer. My research has further benefited from constant support and guidance from Paul Kantor and Dan O'Connor who have always had time for my questions. I want to especially thank Paul for allowing me to work with him during my last semester at Rutgers. I want to thank Zenon Pylyshyn, from the Rutgers Cognitive Science Center, for allowing me to be his research apprentice for several years. Finally, I want to thank Joan Chabrak, Pat Leventhal and Shelee Saal, who have each taken care of me in various ways throughout the years.

I want to thank numerous researchers for their insightful and challenging comments on my dissertation at various workshops, including Marcia Bates, Harry Bruce, Peter Ingwersen, Gary Marchionini, Doug Oard, Dagobert Soergel, Pertti Vakkari, two anonymous reviewers from the User Modeling conference and anyone else who has ever been stuck on an elevator with me at a conference.

I want to thank many of my friends, who also happen to be amazing and exemplar researchers and scientists, for their continued support and friendship over the years: Colleen Cool, Pam Savage-Knepshield, Cynthia Sikora, Patrick Baudisch, David Harper and Jimi Shanahan.

I want to thank the researchers of the future who have assisted and guided me with many of the technical aspects of this dissertation, including Gaurav Sood and Xiaojun Yuan from Rutgers, Fernando Diaz and Haizheng Zhang from the University of Massachusetts, Amherst, and my fellow labmates from the AP Lab, Ying Sun and Ulukbek Ibraev. I want to further thank Vanessa Murdock, from the University of Massachusetts, and Jamie Teevan, from MIT, for their collaborations on other projects. From Rutgers, I would further like to thank Jon Oliver and Sylvia Mueller for their continued support over the years, and Gheorghe Muresan for his help with the parser.

There are many current and former Ph.D. students at SCILS that I would like to thank for their support, encouragement and high tolerance levels: Rebecca Platzner, Sergio Chaparro, Joanne Cattafesta, Jason Delo, Win Quinn, Lynn Cockett-White, Bill White, Shin-jeng Lin and Soo Rieh. I have many friends near and far that I would like to thank for the same reasons: Melissa Beck Lawson, Arlene McBride Joelson, Traci Parnell Ward, Coby Pfeiffer, Helga Dougan, Marlene Gebauer-Camargo, Bill Camargo, Lynn Hickernell, Larisa Donis-Jeppson, Scott Bruton and Louie Barrood.

I want to thank each agency which has provided financial support for this research, and other research that I have done at SCILS: NSF Grant #99-11942, the Eugene Garfield Doctoral Dissertation Fellowship, and the Rutgers Distributed Laboratory for Digital Libraries.

Finally, the seven anonymous subjects who participated in this dissertation research also deserve special recognition and thanks. Their insightfulness, honesty and understanding kept me motivated through the rough spots. One could not find seven better subjects.

DEDICATION

This thesis is dedicated to my parents, Paul and Sue Kelly, and my twin sister, Dawn Kelly Webster (and Woody). Thank you for your love and support.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements.....	iv
Dedication.....	vi
Table of Contents.....	vii
List of Tables.....	xii
List of Illustrations.....	xiv
1. Introduction.....	1
2. Literature Review.....	8
2.1 User Modeling.....	8
2.1.1 User Modeling in Information Retrieval.....	11
2.1.2 Relevance Feedback.....	14
2.1.3 User Modeling and Information Filtering.....	16
2.1.4 Limitations of User Modeling in IR and IF.....	18
2.2 Implicit Methods of Model Acquisition.....	20
2.2.1 Classification of Implicit Feedback Research.....	26
2.2.2 Examination of Implicit Feedback Research.....	30
2.2.2.1 Examine.....	30
2.2.2.2 Retain.....	40
2.2.2.3 Reference.....	41
2.2.2.4 Annotate.....	42
2.2.2.5 Create.....	43

2.3	Limitations of Implicit Modeling Techniques.....	44
2.3.1	Information-seeking Context.....	45
2.3.1.1	Task.....	47
2.3.1.2	Topic.....	52
2.3.2	Concepts of Relevance.....	54
2.3.3	Successive Searching.....	57
2.3.4	Measurement of Behaviors.....	59
2.3.5	Degree of Personalization.....	61
3.	Theoretical Framework.....	64
3.1	Conceptual Model.....	64
3.1.1	General Behavioral Model.....	66
3.1.2	Personalized Models.....	67
3.1.3	Accounting for Information-seeking Context.....	69
3.1.4	Topical Models.....	75
3.2	Research Questions.....	79
3.3	Study Design.....	80
3.3.1	Information-seeking Behaviors.....	81
3.3.2	Information-seeking Context.....	82
3.3.3	Document Preference.....	84
4.	Method.....	86
4.1	Subjects.....	86
4.2	Time Period.....	87
4.3	Instruments and Materials.....	87

4.3.1	Laptops and Printers.....	88
4.3.2	Logging Software.....	89
4.3.3	Consent Form.....	94
4.3.4	Laptop and Printer Agreement.....	94
4.3.5	Entry Questionnaire.....	95
4.3.6	Task and Topic Questionnaires.....	95
4.3.7	Task and Topic Update Questionnaires.....	98
4.3.8	Evaluation Software.....	100
4.3.9	Screen Camera.....	105
4.3.10	Exit Interview.....	106
4.4	Procedure.....	107
5.	Results.....	112
5.1	Characteristics of the Subjects.....	112
5.2	Overview of the Data.....	115
5.3	Tasks.....	117
5.3.1	Endurance.....	122
5.3.2	Frequency.....	126
5.3.3	Stage.....	129
5.4	Topics.....	132
5.4.1	Persistence.....	137
5.4.2	Familiarity.....	140
5.5	Task and Topic Combinations.....	143
5.6	Creating Task and Topic Classes and Classifying Documents.....	144

5.7	Usefulness and Confidence.....	145
5.8	Behaviors and Usefulness.....	149
5.8.1	Display Time.....	149
5.8.2	Retention.....	157
5.9	Behaviors and Information-seeking Context.....	158
5.9.1	Display Time, Task and Topic.....	159
5.9.2	Display Time and Task and Topic Attributes.....	164
5.9.3	Retention, Task and Topic.....	171
5.10	Predicting Usefulness.....	172
5.11	Summary of Results.....	173
6.	Discussion.....	175
6.1	Measuring Information-seeking Context.....	175
6.2	Types of Behaviors that Can Be Gathered.....	181
6.3	Relationships Between Behaviors and Document Preference.....	186
6.3.1	Usefulness Ratings.....	186
6.3.2	Display Time and Usefulness.....	187
6.3.3	Retention and Usefulness.....	190
6.4	Relationship Between Behaviors and Information-seeking Context.....	191
6.4.1	Display Time and Task and Topic.....	191
6.4.2	Retention and Task and Topic.....	193
6.4.3	Display Time and Task and Topic Attributes.....	194
6.5	Predicting Usefulness from Behaviors.....	198
6.6	Relationship of Findings to the Theoretical Model.....	199

7. Conclusions.....	202
Appendix A. Consent Form.....	212
Appendix B. Laptop and Printer Agreement.....	216
Appendix C. Entry Questionnaire.....	217
Appendix D. Task Questionnaire.....	221
Appendix E. Topic Questionnaire.....	222
Appendix F. Task Update Questionnaire.....	223
Appendix G. Topic Update Questionnaire.....	224
Appendix H. Task Update Questionnaire (Version 2).....	225
Appendix I. Topic Update Questionnaire (Version 2).....	226
Appendix J. Web Page Classification.....	227
Appendix K. Exit Interview Schedule.....	229
Appendix L. Task and Topic Combinations.....	230
Appendix M. Natural Display Time According to Usefulness.....	236
Appendix N. Natural Display Time According to Task and Usefulness.....	237
Appendix O. Natural Display Time According to Topic and Usefulness.....	242
Appendix P. Combined Tasks and Excluded Topics from Post-Hoc Analysis.....	249
Appendix Q. Natural Display Time According to Attributes.....	250
References.....	252
Vita.....	261

LIST OF TABLES

2.1 Classification of Behaviors for Implicit Feedback.....23

2.2 Implicit Feedback Research Classified According to Behavior(s) Investigated...28

3.1 Range of Possible Values for Display Time and Topic Familiarity.....72

3.2 Models of Display Time (DT) in Seconds, of Relevant (R) and Non-relevant
(NR) Documents for User A.....74

3.3 Examples of Display Time (DT) in Seconds, Observations and Initiated Action
(BM: Behavioral Model).....75

5.1 Description of Subjects.....113

5.2 Description of the Quantity of Data Collected.....115

5.3 Subject 1’s Tasks.....117

5.4 Subject 2’s Tasks.....118

5.5 Subject 3’s Tasks.....118

5.6 Subject 4’s Tasks.....118

5.7 Subject 5’s Tasks.....119

5.8 Subject 6’s Tasks.....119

5.9 Subject 7’s Tasks.....119

5.10 Subject 1’s Topics.....132

5.11 Subject 2’s Topics.....133

5.12 Subject 3’s Topics.....133

5.13 Subject 4’s Topics.....134

5.14 Subject 5’s Topics.....134

5.15 Subject 6’s Topics.....135

5.16	Subject 7's Topics.....	135
5.17	Descriptive Statistics for Display Time for Each Subject.....	152
5.18	Distributions of Client Display Times According to Percentage of Data.....	153
5.19	ANOVA Results for Display Time and Usefulness.....	155
5.20	Number of Retained Documents (Percent of Total Documents Evaluated).....	157
5.21	ANOVA Results for Display Time and Task and Topic.....	162
5.22	Pair-wise Differences Between Display Times for Task and Topic.....	163
5.23	ANOVA Results for Display Time and Task and Topic Attributes.....	168
5.24	Pair-wise Differences Between Display Times for Task and Topic Attributes...	169
5.25	Tasks and Topics with Retained Documents.....	171

LIST OF ILLUSTRATIONS

3.1	A UM system for personalized interaction.....	66
3.2	Hypothetical behavioral models.....	69
3.3	Behavioral models for endurance of need and familiarity.....	70
3.4	Personalization of interaction.....	73
3.5	Relationship between behavioral models (BM) and topical models (TM).....	77
4.1	Output of the proxy logger (.trec file format).....	93
4.2	Evaluation interface.....	101
5.1	Subjects' experience using computers (a) and computer expertise (b).....	114
5.2	Number of years doing online searching (a) and searching expertise (b).....	114
5.3	Weekly task endurance.....	123
5.4	Frequency of online activities.....	127
5.5	Stage of activities.....	129
5.6	Weekly topic persistence.....	137
5.7	Weekly topic familiarity.....	140
5.8	Mean usefulness of documents evaluated.....	146
5.9	Distributions of usefulness ratings for each subject.....	147
5.10	Mean confidence in usefulness ratings assigned to documents.....	148
5.11	Origin of display time data.....	150
5.12	Distributions of proxy- and client-side display time.....	151
5.13	Distribution of log client display time.....	154
5.14	Mean log display time according to usefulness score.....	155
5.15	Relationship between retention and usefulness.....	158

5.16	Display time according to task and topic for Subject 1.....	159
5.17	Display time according to task and topic for Subject 2.....	160
5.18	Display time according to task and topic for Subject 3.....	160
5.19	Display time according to task and topic for Subject 4.....	160
5.20	Display time according to task and topic for Subject 5.....	161
5.21	Display time according to task and topic for Subject 6.....	161
5.22	Display time according to task and topic for Subject 7.....	161
5.23	Mean log display time and task endurance.....	165
5.24	Mean log display time and frequency.....	165
5.25	Mean log display time and stage.....	166
5.26	Mean log display time and topic persistence.....	166
5.27	Mean log display time and topic familiarity.....	167