

Contents

- 1 Introduction** 1
 - 1.1 Quality of Spoken Dialog Systems 2
 - 1.2 Ensuring that the System will not Fail: Usability Engineering. . . 3
 - 1.3 Automatic Evaluation of Spoken Dialog Systems 5
 - 1.4 Evaluation Versus Learning of the “Optimal Strategy” 6
 - 1.5 Outline and Aim of this Work 7
- 2 MeMo: Usability Workbench** 9
 - 2.1 Model Creation in the MeMo Workbench 11
 - 2.1.1 System Model 11
 - 2.1.2 User Model 12
 - 2.1.3 User Group Editor. 16
 - 2.1.4 Speech Understanding Error Simulation. 16
 - 2.2 Reporting 18
 - 2.2.1 Formative Usability Report 18
 - 2.2.2 Summative Report. 18
 - 2.3 Chapter Summary 19
- 3 Evaluation of the MeMo User Simulation—Use Case Inspire Smart Home System** 21
 - 3.1 Modeling the INSPIRE Smart-Home System in MeMo 22
 - 3.2 Modeling the Experiment 23
 - 3.2.1 Rules for the Simulation 25
 - 3.3 Analysis of Simulated Corpora Using Standard Metrics 26
 - 3.3.1 Results. 27
 - 3.3.2 Comparison to Differences Between Real User Corpora 32

3.4	Manual, Formative Analysis	34
3.4.1	Analysis of User Actions Not Recalled by the Simulation	34
3.4.2	Analysis of Entire Dialogs	37
3.5	Discussion	39
3.6	Chapter Summary	40
4	Detection of Usability Problems Using an Ad-Hoc User Simulation.	41
4.1	Outline of the Study	42
4.1.1	Experimental Data	42
4.1.2	System Model	43
4.1.3	User Models.	44
4.1.4	Speech Understanding Error Model.	48
4.2	Detection of Usability Problems	48
4.2.1	Creating a List of Usability Problems from the Real User Data	49
4.2.2	Classification and Analysis of Problems	51
4.2.3	Problem Discovery in the Simulated Corpora.	52
4.2.4	Preparation of Data for Log File Inspection	55
4.3	Discussion	58
4.4	Chapter Summary	60
5	Prediction of User Judgments.	61
5.1	Data Collection	63
5.1.1	Collection of Quality Issues	64
5.1.2	Selection of the System	64
5.1.3	Conducting the Experiment	66
5.1.4	Analysis of Judgments.	67
5.2	Modeling	68
5.2.1	Performance Measurement	69
5.2.2	Baseline Model with Linear Regression.	71
5.2.3	Modeling the Data with Markov Chains	72
5.2.4	Modeling the Data with Hidden Markov Models	75
5.3	Application to Data with Final Judgments	78
5.3.1	Database	79
5.3.2	Model Training.	80
5.3.3	Results.	80
5.4	Discussion	82
5.5	Chapter Summary	85
6	Application of Prediction Models in a Realistic Usage Scenario . . .	87
6.1	Outline of the Study	88
6.2	Results for Real Data.	90

6.3	Results for Simulated Data	95
6.4	Discussion	98
6.5	Chapter Summary	100
7	Conclusions and Future Work	101
7.1	Evaluation Framework	102
7.2	Modeling of User Judgments and Behavior.	104
7.3	An Integrated Approach	106
Appendix A: List of Usability Problems Found in the BoRIS Database (Chapter 4)		109
Appendix B: List of Usability Issues Covered in the Dialog Scripts of the Experiment in Chapter 5		117
Appendix C: Interaction Parameters Used for Judgment Prediction Models		119
References		123



<http://www.springer.com/978-3-642-31590-9>

Estimating Spoken Dialog System Quality with User Models

Engelbrecht, K.-P.

2013, XIV, 130 p., Hardcover

ISBN: 978-3-642-31590-9