

Table of Contents

Chapter 1. Introduction	1
1.1. INTRODUCTION.....	1
1.2. CONTEXT OF THE STUDY	2
1.3. RESEARCH QUESTIONS	4
1.4. OVERVIEW OF METHOD	4
1.4.1. Think-aloud Triage Studies	5
1.4.2. Identification of Question Taxonomies	6
1.4.3. Classification Task	6
1.5. OVERVIEW OF FINDINGS	8
1.5.1. Think-aloud Triage Studies	8
1.5.2. Classification Task	9
1.6. CONTRIBUTIONS OF THE STUDY	9
1.6.1. Contribution to Theory.....	9
1.6.2. Contribution to Practice	10
1.7. LIMITATIONS OF THE STUDY	11
1.7.1. Focuses on Only One Step in the Process of Providing Digital Reference.....	11
1.7.2. Not System-Building	12
1.8. CHAPTER SUMMARY	12
Chapter 2. Literature Review	13
2.1. INTRODUCTION.....	13
2.2. WHAT IS A QUESTION?.....	13
2.2.1. Questions from the Perspective of Erotetic Logic	14
2.2.2. Questions from the Perspective of Speech Act Theory	15
2.2.3. Questions from the Perspective of Discourse Analysis.....	17
2.2.4. Questions from the Perspective of Information Needs and Uses	20
2.2.5. Defining the Term “Question”	22
2.3. REFERENCE.....	24
2.3.1. Information Needs	26
2.3.2. The Reference Interview and Question Negotiation	27
2.4. KNOWLEDGE-BASED SYSTEMS FOR REFERENCE	28
2.4.1. What is a Knowledge-based System?	29
2.4.2. The Functionality of Knowledge-based Systems for Reference.....	30
2.4.3. The Failure of Knowledge-based Systems for Reference ..	32
2.5. DIGITAL REFERENCE	33
2.5.1. A Model of Digital Reference	36

2.5.2. The Triage Process	37
2.5.2.1. Variations on the Triage Process.....	40
2.5.2.2. Triage in Real-Time Services	41
2.5.3. The Opening Question in Digital Reference	42
2.5.4. Digital Reference vs. Knowledge-based Systems for Reference	44
2.6. QUESTION ANSWERING SYSTEMS	46
2.6.1. QUALM	46
2.6.2. The TREC Question Answering Track	47
2.6.3. The Possible Future of Question Answering Systems	50
2.7. CLASSIFICATION	53
2.7.1. Classification and Categorization	53
2.7.2. Classification of Intellectual Entities	54
2.7.3. Mutual Exclusivity	54
2.7.4. Automatic Classification and Automatic Indexing	55
2.7.5. Faceted Classification Schemes	57
2.7.6. Evaluation of Classification Schemes	59
2.8. SURVEY OF EXISTING QUESTION TAXONOMIES	62
2.8.1. Subjects of Questions	64
2.8.2. Wh- Words	66
2.8.3. The Functions of Expected Answers	68
2.8.4. The Forms of Expected Answers	71
2.8.4.1. The Subject of the Question	71
2.8.4.2. Types of Sources from Which the Answer May Be Drawn	72
2.8.4.3. The Forms of Expected Answers.....	73
2.8.5. Linguistic Analysis	76
2.9. CHAPTER SUMMARY	77
Chapter 3. Methodology	79
3.1. INTRODUCTION	79
3.1.1. Research Questions	80
3.1.2. Research Design	81
3.2. PHASE 1: TRIAGE THINK-ALLOUD STUDIES	85
3.2.1. The Think-aloud Method.....	85
3.2.2. The Constant Comparative Method.....	89
3.2.3. Selection of Digital Reference Services	91
3.2.4. Solicitation of Respondents	95
3.2.5. Think-aloud Task	96
3.2.5.1. Prompts and Probes.....	97
3.2.5.2. Content Analysis and the Unit of Analysis	99
3.2.6. Post-Think-aloud Survey	102
3.2.7. Coding of Reasons for Triage Decisions	103

4.3.1.1.	Taxonomy of Wh- Words	175
4.3.1.1.1.	Intersection of the Taxonomies of Wh- Words and Functions of Expected Answers	177
4.3.1.1.2.	Intersection of the Taxonomies of Wh- Words and Forms of Expected Answers.....	178
4.3.1.2.	Taxonomy of Functions of Expected Answers.....	179
4.3.1.2.1.	Intersection of the Taxonomies of Functions of Expected Answers and Wh- Words.....	181
4.3.1.3.	Taxonomy of Forms of Expected Answers	182
4.3.1.3.1.	Intersection of the Taxonomies of Forms of Expected Answers and Wh- Words.....	184
4.3.1.3.2.	Intersection of the Taxonomies of Forms of Expected Answers and Functions of Expected Answers.....	185
4.3.1.4.	The Intersection of All Three Taxonomies	186
4.3.2.	Intercoder Reliability	187
4.3.3.	Evaluation of Question Taxonomies.....	190
4.3.3.1.	Evaluation of the Taxonomy of Wh- Words	190
4.3.3.2.	Evaluation of the Taxonomy of Functions of Expected Answers	192
4.3.3.3.	Evaluation of the Taxonomy of Forms of Expected Answers ..	193
4.3.3.4.	Summary of the Evaluation	194
4.3.4.	Modification of Question Taxonomies	196
4.3.4.1.	Modifications to the Taxonomy of Wh- Words	196
4.3.4.2.	Modifications to the Taxonomy of Functions of Expected Answers	197
4.3.4.3.	Modifications to the Taxonomy of Forms of Expected Answers	199
4.4.	PHASE 3: CLASSIFICATION TASK.....	200
4.4.1.	Classification of Questions	200
4.4.1.1.	Taxonomy of Wh- Words	201
4.4.1.2.	Taxonomy of Functions of Expected Answers.....	203
4.4.1.3.	Taxonomy of Forms of Expected Answers	205
4.4.2.	Unclassifiable Questions	206
4.4.3.	Questions on which Coders Disagreed	208
4.4.3.1.	Taxonomy of Wh- Words	209
4.4.3.2.	Taxonomy of Functions of Expected Answers.....	210
4.4.3.3.	Taxonomy of Forms of Expected Answers	210
4.4.4.	Intercoder Reliability	212
4.4.5.	Correlation between Question Type and Triage Action.....	213
4.5.	CHAPTER SUMMARY	220

Chapter 5. Discussion222

5.1.	INTRODUCTION.....	222
5.1.1.	Defining the Term “Question” for Triage.....	223

5.2. THE UTILITY OF QUESTION TAXONOMIES.....	224
5.2.1. Linguistic Analysis	225
5.2.2. Question Answering.....	226
5.2.3. Digital Reference	227
5.2.4. Faceted Classification.....	228
5.2.5. Other Possible Taxonomies.....	229
5.3. ATTRIBUTES THAT AFFECT THE TRIAGE PROCESS	232
5.3.1. Training for Digital Reference Triagers.....	232
5.3.2. Machine-Assisted Triage	234
5.3.3. Grammars of Actions	238
5.3.4. The Equivalence of Triage and Answer Formulation.....	239
5.3.5. Correctness of the Triage Decision	241
5.3.6. The Zipfian Distribution.....	243
5.4. THE UNIQUENESS OF THE TRIAGE PROCESS	246
5.5. TRIAGE IS DEAD, LONG LIVE TRIAGE.....	248
5.6. THE QUESTION INTERCHANGE PROFILE (QUIP)	251
5.7. REAL-TIME REFERENCE	252
5.8. CHAPTER SUMMARY	254
Appendix A: Solicitation Email Messages.....	261
Appendix B: Think-Aloud Instructions	265
Appendix C: Taxonomies, Scope Notes, and Coder Instructions Utilized in Phase 2	267
Appendix D: Taxonomies, Scope Notes, and Coder Instructions Utilized in Phase 3	274
Appendix E: Questions Coded in Phase 3	284
Bibliography.....	286

Table of Tables

Table 2-1: An Overview of Perspectives on Questions.....	22
Table 2-2: Types of Systems and Their Approach to Question Answering.....	52
Table 2-3: Thirteen Criteria for the Evaluation of Classification Schemes	61
Table 2-4: The Taxonomy of Functions of Expected Answers	69
Table 2-5: The Taxonomy of Forms of Expected Answers	75
Table 2-6: Levels of Linguistic Analysis of Question Taxonomies	77
Table 3-1: Steps in the Phase 1 Methodology: Think-aloud Triage Studies	82
Table 3-2: Steps in the in the Phase 2 Methodology: Identification of Question Taxonomies	83
Table 3-3: Steps in the Phase 3 Methodology: Classification Task	84
Table 3-4: Thirteen Criteria for the Evaluation of Classification Schemes	112
Table 3-5: Classification of Sets of Questions among Groups of Coders	118
Table 4-1: Reasons for Services' Not Participating in the Study.....	128
Table 4-2: Number of Think-aloud Questions Received on the Day the Think-aloud Study was Performed	138
Table 4-3: Categories and Attributes That Affect Triagers' Decisions	141
Table 4-4: Attributes that Affect Triage Decisions.....	148
Table 4-5: Questions Triageed by Attribute Category.....	152
Table 4-6: Break-down of Triage Destinations of Questions from High- and Low-Volume Services	162
Table 4-7: Break-down of Triage Destinations of Questions from High- and Low-Volume Services	164
Table 4-8: Post-Think-aloud Survey Results	169
Table 4-9: Questions Classified According to the Taxonomy of Wh- Words	176
Table 4-10: Questions Classified According to the Taxonomy of Functions of Expected Answers	179
Table 4-11: Questions Classified According to the Taxonomy of Forms of Expected Answers	182
Table 4-12: Phase 2 Intercoder Reliability Kappa Values.....	188
Table 4-13: Thirteen Criteria for the Evaluation of Classification Schemes	195
Table 4-14: Questions Classified According to the Taxonomy of Wh- Words ..	201
Table 4-15: Questions Classified According to the Taxonomy of Functions of Expected Answers	203
Table 4-16: Questions Classified According to the Taxonomy of Forms of Expected Answers	205
Table 4-17: Phase 3 Intercoder Reliability Kappa Values.....	213
Table 4-18: Values of Cramér's V for the Correlation between Question Taxonomies and Question Attributes.....	214
Table 4-19: Values of Cramér's V for the Correlation between Question Taxonomies and Triage Action	215

Table 4-20: Values of Cramér’s V for the Correlation between Question Taxonomies and Triage Action, in the Four Highest-volume Services	216
Table 4-21: Very Strong Correlations between Intersections of Question Types and Triage Action	218
Table 5-1: Question Taxonomies and Levels of Linguistic Analysis	226
Table 5-2: Categories and Attributes That Affect Triagers’ Decisions	254
Table 5-3: Very Strong Correlations between Intersections of Question Types and Triage Action	257

Table of Figures

Figure 1-1: General Digital Reference Model	3
Figure 1-2: Research Methodology	5
Figure 2-1: The Process of Semantic Collaboration (Roberts and Bavelas, 1996, p. 143).....	19
Figure 2-2: Steps in the Triage Process	40
Figure 3-1: Research Methodology	81
Figure 4-1: Research Methodology	125
Figure 4-2: Distribution of Number of Questions Received by Participating Services on the Day the Think-aloud Study was Performed	139
Figure 4-3: Distribution of Question Attributes that Affect Triage Decisions	150
Figure 4-4: Distribution of all Attributes that Affect Triage Decisions	151
Figure 4-5: Distribution of Questions Triageed by Attribute Category.....	152
Figure 4-6: Distribution of all Attributes that Affect Triage Decisions, with the 15 Factors from Pomerantz, Nicholson, and Lankes (2003) Highlighted.....	156
Figure 4-7A: Distribution of Question Attributes that Affect Triage Decisions, in total	166
Figure 4-7B: Distribution of Question Attributes that Affect Triage Decisions, in total	167
Figure 4-7C: Distribution of Question Attributes that Affect Triage Decisions, in total	168
Figure 4-8: General Model of the Triage Process.....	172
Figure 4-9: Distribution of Questions Classified According to the Taxonomy of Wh- Words.....	176
Figure 4-10: Distribution of Questions Classified According to the Taxonomy of Functions of Expected Answers.....	180
Figure 4-11: Distribution of Questions Classified According to the Taxonomy of Forms of Expected Answers	183
Figure 4-12: Distribution of Questions Classified According to the Taxonomy of Wh- Words.....	202
Figure 4-13: Distribution of Questions Classified According to the Taxonomy of Functions of Expected Answers.....	204
Figure 4-14: Distribution of Questions Classified According to the Taxonomy of Forms of Expected Answers	206
Figure 4-15: Taxonomy Space Defined by the Three Taxonomies.....	217

Figure 5-1: Taxonomy Space as the Basis for Machine-Assisted Triage	236
Figure 5-2: Screenshot of the Triage Area for a Digital Reference Service Using the QABuilder Software	249