Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://dare.uva.nl/document/191345

File ID	191345
Filename	Contents

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	Dissertation
Title	Search in audiovisual broadcast archives
Author	B. Huurnink
Faculty	Faculty of Science
Year	2010
Pages	viii, 198
ISBN	978-90-786-7599-0

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/358972

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Contents

Acknowledgments **vii**

1 Introduction **1**
1.1 Research Questions 3
1.2 Organization 5
1.3 Main Contributions 6
1.4 Origins of the Thesis 7

I. Searchers in Audiovisual Broadcast Archives **8**

2 Audiovisual Searchers **11**
2.1 User Studies in Information Retrieval 11
2.2 Field Studies of Audiovisual Retrieval 13
2.2.1 Web Searchers 13
2.2.2 Archive Users 14
2.3 Laboratory Experiments 15
2.3.1 Experiments Involving Students and Researchers 16
2.3.2 Professionals in the Lab 17
2.3.3 Simulated Searchers 18
2.4 Discussion 19

3 Search Behavior of Media Professionals **21**
3.1 Organizational Setting 22
3.1.1 Users of the Archive 23
3.1.2 Finding and Ordering Audiovisual Material 23
3.1.3 Audiovisual Thesaurus 26

3.2 Experimental Design . . . 27
3.2.1 Data Collection . . . 27
3.2.2 Definitions . . . 28
3.3 Transaction Log Analysis . . . 29
3.3.1 Session-level Analysis . . . 30
3.3.2 Query-level Analysis . . . 33
3.3.3 Term-level Analysis . . . 35
3.3.4 Facet-level Analysis . . . 38
3.3.5 Order-level Analysis . . . 39
3.4 Conclusion . . . 42
3.4.1 Answers to Research Questions . . . 43
3.4.2 Implications for the Thesis . . . 44

4 Simulating Logged User Queries 45
4.1 Related Work . . . 47
4.2 Simulation Framework . . . 49
4.2.1 Validating Simulators . . . 49
4.2.2 Simulating Purchase Queries . . . 50
4.3 Experimental Setup . . . 54
4.3.1 Experimental Data . . . 55
4.3.2 Simulator Settings . . . 55
4.3.3 Retrieval Systems . . . 55
4.4 Results and Analysis . . . 56
4.5 Conclusion . . . 58

5 Conclusion to Part I 61

II. Content-Based Video Retrieval for Audiovisual Broadcast Archives 63

6 Methods for Content-Based Video Retrieval 65
6.1 The Role of Benchmarking . . . 66
6.2 Searching with Automatically Generated Content Metadata . . . 69
6.2.1 Transcript-Based Search . . . 70
6.2.2 Low-Level Feature-Based Search . . . 71
6.2.3 Detector-Based Search . . . 72
6.2.4 Combining Results from Multiple Methods . . . 74
6.3 Mismatches in Content-Based Video Retrieval . . . 75
6.3.1 The Vocabulary Mismatch when Selecting Concepts . . . 76
6.3.2 The Temporal Mismatch Between the Audio and Video Signals 78

6.4 Conclusions 79

7 Two Collections 81
7.1 The Broadcast News Collection 81
7.1.1 Collection Overview 82
7.1.2 A Unified Multimedia Thesaurus 83
7.2 The Archive Footage Collection 85
7.2.1 Collection Overview 85
7.2.2 Archive Query Set 86
7.2.3 Future Query Set 88
7.2.4 Lab Query Set 90
7.3 Conclusion 92

8 Assessing Concept Selection for Video Retrieval 93
8.1 Concept Selection Benchmark Development 95
8.1.1 Collection Benchmark 96
8.1.2 Human Benchmark 98
8.2 Analysis of the Concept Selection Benchmarks 100
8.2.1 Semantic Overlap Between the Benchmarks 100
8.2.2 Retrieval Performance 104
8.2.3 Reproducibility 108
8.3 Using the Benchmarks to Assess Concept Selection 110
8.3.1 Assessment Measures 110
8.3.2 Test Case: Assessing Concept Selection 111
8.4 Discussion 115
8.4.1 Analyzing the Benchmarks 115
8.4.2 Evaluating Concept Selection Algorithms 116
8.5 Conclusions 116

9 Redundancy and the Temporal Mismatch 119
9.1 Discovering Redundancy 122
9.1.1 Methodology 122
9.1.2 Experimental Setup 123
9.1.3 Redundancy in the Video Signal 124
9.1.4 Redundancy Across the Video and Audio Signals 126
9.1.5 Estimating Expected Visual Relevance 127
9.2 Retrieval Framework 128
9.2.1 Retrieval Based on Language Modeling 128
9.2.2 Document Expansion 129
9.2.3 Integrating Redundancy 130

9.3 Retrieval Experiments 131
9.3.1 Result Overview 131
9.3.2 Impact at the Item Level 133
9.3.3 Comparing the redundancy models 135
9.4 Conclusions and Future Work 137

10 Integrating Content-Based Video Retrieval into the Archive 139
10.1 Evaluation Methodology 141
10.1.1 Query Definitions 141
10.1.2 Retrieval Data Sources 143
10.1.3 Video Retrieval Tasks 143
10.1.4 Video Search Engines 143
10.2 Experimental Setup 145
10.3 Results 146
10.3.1 Experiment 1: 3x3 Shot Retrieval 146
10.3.2 Experiment 2: 3x3 Program Retrieval 148
10.3.3 Experiment 3: Program Retrieval with Simulated Queries . . . 148
10.3.4 Experiment 4: Prioritizing Content Search 149
10.4 Conclusions and Recommendations 149

11 Conclusion to Part II 153

12 Conclusions to the Thesis 155
12.1 Answers to Research Questions 155
12.2 Main Contributions 157
12.3 Future Directions 159

Bibliography 161

A Identifying Title and Thesaurus Terms in Queries 175

B Visual Concepts in the Multimedia Thesaurus 179

C Queries 183

Samenvatting 187

Summary 189