UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

System evaluation of archival description and access

Zhang, J.

Publication date
2011

Link to publication

Citation for published version (APA):
Zhang, J. (2011). *System evaluation of archival description and access*. [Thesis, fully internal, Universiteit van Amsterdam]. IR Publications.

General rights
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, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

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

Download date:18 Nov 2026

Contents

Preface **ix**

1 Introduction **1**
1.1 The Research Work 2
1.1.1 Problem Statement 2
1.1.2 Scope and Terminology 3
1.1.3 Research Plan and Contributions 4
1.1.4 Research Questions 4
1.2 Structure of the Thesis 5

2 Usage of XML Retrieval for Archival Access **7**
2.1 Introduction 7
2.2 Related Work 8
2.2.1 Extensible Markup Language 9
2.2.2 Encoded Archival Description 10
2.2.3 XML Information Retrieval 11
2.3 Archival Access: Information Retrieval with EAD 16
2.3.1 Retrieval with EAD 16
2.3.2 Archival Metadata Retrieval Systems 18
2.4 README System Description 19
2.4.1 Usage Scenarios 19
2.4.2 EAD Finding Aids 21
2.4.3 Indexing and Storage 22
2.4.4 Retrieval Model 23
2.4.5 Querying and User Interfaces 24

2.4.6 Result Delivery 30
2.5 Conclusion 30

3 Construction of an Archival Test Collection for Evaluation 33
3.1 Introduction 33
3.2 Background and Related Work 35
3.2.1 Log Analysis for Information Retrieval Evaluation 35
3.2.2 IR Test Collections 36
3.2.3 Evaluating Archival Metadata Retrieval 36
3.3 From Search Log to Test Collection 37
3.3.1 *Nationaal Archief* System Interaction 37
3.3.2 Search Log Files and Document Collection 41
3.3.3 Information Extraction from Logs 41
3.3.4 Types of Test Collections 42
3.4 Log-Based Evaluation in Action 43
3.4.1 IR Models and Systems 43
3.4.2 Study 1: Complete Log Test Collection 44
3.4.3 Study 2: Test Collection Based on Agreements 45
3.5 External Validation 47
3.5.1 Test Collection Based on E-Mail Reference Requests . . . 47
3.5.2 System Rank Correlations 47
3.5.3 Comparison with the *Nationaal Archief* System 49
3.5.4 The War Research Guide 50
3.6 Conclusion 54

4 On Archival Description Principles for Retrieval 57
4.1 Introduction 57
4.2 Problem Statement 59
4.3 Test Collection for Within-Fonds Retrieval 60
4.3.1 XML IR Test Collection with Manual Assessments 60
4.3.2 From Web Anchor to XML Element 63
4.4 The Scope of EAD Markup for Retrieval 65
4.4.1 Background 65
4.4.2 Analysis and Results 65
4.5 Retrieval Effects of Grouping Descriptions 67

4.5.1 Background . . . 67
4.5.2 Study 1: Measuring Concentration Ratios . . . 69
4.5.3 Study 2: Retrieval Effects of Provenance . . . 72
4.6 Original Order Versus Relevance . . . 73
4.6.1 Background . . . 74
4.6.2 Study 1: Using Manual Assessments Within a Fonds . . . 75
4.6.3 Study 2: Experiment with Log-Based Approach . . . 77
4.7 Discussion . . . 78
4.8 Conclusion . . . 79

5 Searching Within EAD Finding Aids 81
5.1 Introduction . . . 81
5.2 Related Work . . . 83
5.2.1 Deriving Web Search Behavior from Search Logs . . . 83
5.2.2 Search Behavior in Cultural Heritage . . . 85
5.3 Analysis of Searching Within an EAD Finding Aid . . . 86
5.3.1 A First Exploration . . . 86
5.3.2 The States of Within-EAD Finding Aid Search . . . 87
5.3.3 Within-EAD Finding Aid Search Transition Diagram . . . 90
5.4 Understanding Archival Search Behaviors . . . 91
5.4.1 User Groups and Variables . . . 91
5.4.2 Study 1: Pioneering into Finding Aids . . . 93
5.4.3 Study 2: Incremental Interaction and Search Behavior . . . 99
5.4.4 Study 3: On the Uniformity of Archival Search . . . 104
5.5 Conclusion . . . 109

6 User Stereotypes and Evaluation 111
6.1 Introduction . . . 111
6.2 Related Work . . . 113
6.2.1 Transaction Log Analysis . . . 113
6.2.2 User Stereotypes and Evaluation . . . 114
6.2.3 Improving IR Systems and Search Behavior . . . 115
6.3 Archival Search Log Analysis . . . 117
6.3.1 Preparation . . . 117
6.3.2 Analysis . . . 117

6.4 Deriving User Groups 122
6.4.1 Implicit Features of User Interest 122
6.4.2 Vocabulary Use 123
6.4.3 Results 123
6.4.4 Novices and Experts 127
6.4.5 Correlations 128
6.5 IR Evaluation in Context 129
6.5.1 Experimental Setup 129
6.5.2 Filtering Assessments from User Groups 130
6.5.3 Results 130
6.6 Conclusion 133

7 Conclusions **135**
7.1 Findings and Contributions 135
7.1.1 Main Findings and Contributions 135
7.1.2 General Contributions and Findings 136
7.1.3 IR and Archival Science Contributions 142
7.2 Future Work 142

Bibliography **143**

Abstract **155**

Samenvatting **157**

Titles in the ILLC Dissertation Series **159**

Titles in the SIKS Dissertation Series **161**