Contents

Contents I

List of Abbreviations VI

1 Introduction 1

1.1 Motivation 1

1.2 Research Design and Research Questions 3

1.3 Thesis Outline 4

2 Definition of Key Terms and Methods 7

2.1 Introduction 7

2.2 Definition of Key Terms 7

2.2.1 Job Requirements 7

2.2.2 Online Job Advertisements 8

2.3 Methods of Job Requirement Analysis 11

2.3.1 Quantitative Methods in Social Research 11

2.3.2 Job Advertisement Analysis 13

2.4 Summary 15

3 State of the Art in Technology of Web Content Mining 17

3.1 Introduction 17

3.2 Definitions and Related Research Fields 17

3.2.1 Terms and Definitions 17

3.2.2 Related Research Fields 22

3.3 Motivational Background 24

3.4 Methodological Foundations of Web Content Mining 26

3.4.1 Overview 26

3.4.2 The KDD Process 26

3.4.3 Subtasks of Web Mining 29

Bibliografische Informationen
http://d-nb.info/1079990658

digitalisiert durch DEUTSCHE NATIONAL BIBLIOTHEK

3.4.4 Summary 30

3.5 Recent Research on Web Content Mining 31

3.5.1 Overview 31

3.5.2 Data Sources and Search Procedure 31

3.5.3 Overview of Previous Review Articles and Text Books 32

3.5.4 Overview of Relevant Research Articles on Web Content Mining 34

3.5.5 Summary 38

3.6 Summary 38

4 Comparative Analysis of Methods for Job Requirement Analysis 41

4.1 Introduction 41

4.2 Surveys using Quantitative Methods from Social Research 41

4.2.1 Quantitative Surveys using Questionnaires 41

4.2.2 Quantitative Surveys using Structured Interviews 44

4.2.3 Quantitative Surveys using Both Methods 45

4.3 Surveys using Job Advertisement Analysis 47

4.3.1 Job Advertisement Analyses using Content Analysis 47

4.3.2 Job Advertisement Analyses using Big Data Analytics 51

4.4 Comparative Analysis of Survey Methods 54

4.4.1 Introduction 54

4.4.2 Purpose and Scope of Surveys 54

4.4.3 Applied Methods 55

4.4.4 Data Sources and Characteristics 57

4.4.5 Sampling 59

4.4.6 Summary 66

4.5 Conclusion 68

5 Design and Implementation of the Web Content Mining Process 71

5.1 Introduction 71

5.2 Design of the Web Content Mining Process 71

5.3 Implementation of Web Information Retrieval and Search 75

5.3.1 Terms and Definitions 75

5.3.2 Web Crawling 80
5.3.3 Characteristics of Target Data 83
5.3.4 Problems and Solutions 84

5.4 Implementation of Information Selection and Preprocessing 85
5.4.1 Terms and Definitions 85
5.4.2 Data Cleaning 90
5.4.3 Language Detection 92
5.4.4 Natural Language Processing 95
5.4.5 Characteristics of Preprocessed Data 96
5.4.6 Problems and Solutions 96

5.5 Implementation of Information Extraction and Integration 97
5.5.1 Terms and Definitions 97
5.5.2 Web Information Extraction 99
5.5.3 Integration 100
5.5.4 Characteristics of Transformed Data 100
5.5.5 Problems and Solutions 101

5.6 Implementation of Generalization 101
5.6.1 Terms and Definitions 101
5.6.2 N-Grams and Word Co-Occurrences 103
5.6.3 Characteristics of Patterns 104
5.6.4 Problems and Solutions 104

5.7 Implementation of Analysis and Testing of Mined Patterns 105
5.7.1 Terms and Definitions 105
5.7.2 Interpretation 106
5.7.3 Validation 107
5.7.4 Problems and Solutions 109

5.8 Implementation of Visualization 110
5.8.1 Terms and Definitions 110
5.8.2 Visual Representations 110
5.8.3 Problems and Solutions 112

5.9 Summary 112

6 Results of Applying the Web Content Mining Process 115

6.1 Introduction 115

6.2 General Description of the Sample 115

6.3 References to Professional Requirements 118

6.3.1 Qualification Level 118

6.3.2 Professional Experiences 122

6.3.3 Language Requirements 126

6.4 References to Methodological Requirements 128

6.5 References to Social Requirements 130

6.6 References to Personal Requirements 132

6.7 References to the Job Position 136

6.8 References to the Job Task 138

6.9 References to the Company 142

6.9.1 Location 142

6.9.2 Business Sectors 147

6.9.3 Internationalization 150

6.10 Comparison with the Results of Previous Surveys on Job Requirements 152

6.10.1 Introduction 152

6.10.2 Professional Requirements 153

6.10.3 Methodological Requirements 154

6.10.4 Social Requirements 155

6.10.5 Personal Requirements 155

6.10.6 Job Positions & Job Tasks 156

6.10.7 Company 156

6.11 Conclusion 157

7 Evaluation of the Web Content Mining Process 159

7.1 Introduction 159

7.2 Quality Criteria for Research 159

7.3 Evaluation Procedure 160

7.4 Indicators 161

7.5 Objectivity 166

7.6 Validity 167

7.7 Reliability 168

7.8 Summary 169

8 Conclusion and Future Research 171

8.1 General Results 171

8.2 Limitations 173

8.3 Transfer 175

8.4 Future Research 178

9 Summary 183

10 References 185

10.1 List of Figures 185

10.2 List of Tables 189

10.3 List of Equations 191

10.4 Bibliography 193

11 Annex 205

11.1 Overview of Relevant Research Articles on Web Content Mining 205

11.2 Overview of Surveys using Quantitative Methods from Social Research 206

11.3 Overview of Job Advertisement Analyses using Content Analysis 212

11.4 Overview of Job Advertisement Analyses using Big Data Analytics 218

11.5 XML Code of N-Gram Analysis 222

11.6 Term Occurrences Related to Job Tasks 225

11.7 Term Occurrences Related to Location 228