Contents

List of Figures XIII
List of Tables XVI
List of Abbreviations XIX

Introduction **1**

1 Web Search **5**
1.1 Information Demand 7
1.1.1 User's Information Need 7
1.1.2 User Behavior 10
1.2 Search Tools 13
1.3 How Search Engines Work 19
1.3.1 Crawler 20
1.3.2 Indexer 33
1.3.3 Analyzer 35
1.3.4 Searcher 36
1.3.5 Presenter 37
1.4 A New Approach to Relevancy 38
1.5 Evolution of Web Search Engines 46

2 Web Structure **49**
2.1 Evolution of Information Supply 49
2.2 Web Content 50
2.3 The Web Graph 53
2.3.1 Reasons for Regarding the Web Graph 54
2.3.2 Properties of the Web Graph 55
2.3.3 Representation and Storage of the Web Graph 61
2.4 Link Analysis 67
2.4.1 Ranking 67
2.4.2 Authorities and Hubs 71
2.4.3 Other Link-based Methods 77
2.5 Data Structures 83
2.5.1 Feasible Data Structures 84

2.5.2 Evaluation . . . 89
2.5.3 Selection . . . 93
2.6 HITS in Practice . . . 95
2.6.1 Implementation of the Algorithm . . . 95
2.6.2 Experimental Results and Evaluation . . . 100

3 Result Clustering **105**
3.1 Clustering Methods . . . 106
3.2 Web Clustering . . . 108
3.3 Text Based Approaches . . . 109
3.4 Link Based Approaches . . . 112
3.4.1 Robust Clustering Algorithm . . . 113
3.4.2 Inclusion of Link Direction . . . 119
3.4.3 Architecture . . . 121
3.4.4 Experimental Results . . . 123
3.5 Combined Approaches . . . 125
3.5.1 Architecture . . . 125
3.5.2 Preprocessing . . . 126
3.5.3 Document Representation . . . 130
3.5.4 Clustering Method Combination . . . 133
3.5.5 Implementation . . . 151
3.5.6 Experimental Results . . . 153

4 Search Engine Optimization **163**
4.1 Objectives . . . 163
4.1.1 Website Visitors . . . 166
4.1.2 Result Page Position . . . 168
4.2 Possibilities and Means . . . 171
4.2.1 Visibility . . . 173
4.2.2 Static Relevance . . . 174
4.2.3 Dynamic Relevance . . . 175
4.3 Keyword Selection . . . 176
4.3.1 Keyword Selection Position Model (KSPM) . . . 178
4.3.2 Enhanced Keyword Selection Position Model (EKSPM) . . . 182
4.3.3 Keyword Selection Top Position Model (KSTPM) . . . 184
4.4 Limits . . . 185
4.4.1 Search Engine Spam . . . 186
4.4.2 Spamming Methods . . . 187
4.5 Implementation . . . 188

4.5.1 Tasks 189
4.5.2 Website Creation Process 191
4.6 Practical Results 195
4.6.1 Project and Test Websites 195
4.6.2 Reporting and Results 197

5 Conclusion **205**

Bibliography **207**

Index **219**