

Contents

1	Introduction	1
1	From Archie to Google and Beyond	1
2	The Academic and Industrial Perspectives	1
3	Paradigm Shifts	2
4	A Robust Retrieval Model	5
5	Outline	6
2	Classical Retrieval Models	7
1	Overview	7
2	Bag of Words Models	7
2.1	Binary Independence Retrieval Model	8
2.2	2-Poisson Model	12
2.3	BM25 Model	13
2.4	Unigram Language Models	14
2.5	Other Bag of Words Models	15
3	Models That Go Beyond the Bag of Words Assumption	16
3.1	n -Gram Language Models	16
3.2	Indri Inference Network Model	17
3.3	Other Models That Go Beyond the Bag of Words Assumption	21
4	The Current State-of-the-Art	22
3	Feature-Based Ranking	23
1	Overview	23
2	Modeling Relevance	23
3	The Markov Random Field Model	24
3.1	Graph Structure	26
3.2	Potential Functions	27
4	Constructing Markov Random Fields	29
4.1	Dependence Model Type	30
4.2	Clique Set Type	30

4.3	Weighting Function	32
4.4	Examples	35
5	Ranking	37
6	Ad Hoc Retrieval	39
6.1	MRF Models for Ad Hoc Retrieval	41
6.2	Evaluation	49
6.3	Summary of Results	66
7	Web Search	67
7.1	Previous Models for Web Search	69
7.2	Document Priors	70
7.3	MRF Models for Web Search	73
7.4	Results	75
4	Feature-Based Query Expansion	79
1	Overview	79
2	Related Work	80
3	Basic Latent Concept Expansion	81
3.1	Query Expansion	84
3.2	Comparison to Relevance Models	84
4	Generalized Latent Concept Expansion	85
4.1	Features	86
4.2	Comparison with Previous Models	86
5	LCE Using Hierarchical MRFs	88
5.1	Data Representation	88
5.2	Model Description	89
5.3	Discussion	91
5.4	Time Complexity	92
5.5	HMRFs for Ranking	93
6	Experimental Results	94
6.1	Ad Hoc Retrieval Results	94
6.2	Robustness	97
6.3	Multi-term Concept Generation	98
6.4	Evaluation of LCE-GE and LCE-HMRF	100
7	Discussion	105
7.1	Relevance vs. Relevant Documents	105
7.2	The Role of Dependence	105
5	Query-Dependent Feature Weighting	107
1	Overview	107
2	Related Work	108
3	Weighted Dependence Model	109
4	Concept Importance Features	111
5	Evaluation	113
5.1	Experimental Setup	113
5.2	TREC Evaluation	114
5.3	Large-Scale Web Evaluation	117

- 6 Model Learning 121**
 - 1 Parameter Estimation 121
 - 1.1 Direct Search 122
 - 1.2 Optimization Using Surrogate Functions 125
 - 2 Feature Selection 128
 - 2.1 Related Work 129
 - 2.2 Automatic Feature Selection 129
 - 2.3 Evaluation 131
 - 3 Learning to Efficiently Rank 136
 - 3.1 Related Work 137
 - 3.2 Tradeoff Metrics 138
 - 3.3 Model 141
 - 3.4 Parameter Estimation 143
 - 3.5 Experimental Results 143
- Appendix A Data Sets 149**
 - 1 Anatomy of a TREC Data Set 149
 - 2 Summary of Data Sets 151
- Appendix B Evaluation Metrics 153**
- References 157**
- Index 165**