

Contents

Part I Understanding the Blogosphere

1	Introduction: The Blogosphere	3
1.1	Origins of Social Network Research	3
1.2	From the Web to the Social Web	4
1.3	Dimensions of Web 2.0	4
1.4	Web 2.0 and Weblogs	7
1.5	Publishing Revolution or Utopian Fallacy	9
1.6	Weblogs vs Facebook, Twitter & Co.	10
2	Micro-perspective	13
2.1	Weblogs: The Smallest Entities of the Blogosphere	13
2.2	Blogging Software and Platforms	13
2.3	Hosting Issues of Blogging	15
2.4	Weblog Features	17
2.5	Blog Typology: Classification	18
2.5.1	Classification by Genre/Content	18
2.5.2	Classification by Author(s)	19
2.5.3	Classification by Information Flow	22
2.5.4	Hybrid Forms of Classification	22
2.5.5	Weblog Typology: Conclusion	24
3	Macro-perspective	25
3.1	Social Software	25
3.2	Social Physics	26
3.2.1	Power Laws	27
3.2.2	The Pareto or Zipfian Distribution	28
3.2.3	The Long Tail of Blogging	28
3.2.4	Social Physics	29

3.3	Disruptive Technologies: Changing the Rules of the Game.....	31
3.3.1	Failure Framework by Christensen	32
3.3.2	The Disruptive Character of the Web 2.0 and Weblogs...	34

Part II The Continents of the Blogosphere

4	Overview of the Continents of the Blogosphere	37
4.1	Docu-Blogs	37
4.2	Edu-Blogs	38
4.3	Ego-Blogs	38
4.4	Corporate Blogs	38
5	Continent of Docu-Blogs Use Case: The IT-Gipfelblog.....	39
5.1	Politics and the Blogosphere.....	40
5.2	The IT-Summit Series	42
5.3	Blogs: Means of Expression for Direct Democratic Politics	43
5.4	Quality Management and User Control	43
5.5	Continuous Adaptation.....	44
5.6	Typologisation of a Docu-Blog	45
5.7	The Contentual Development.....	46
5.8	The Influence of External Factors on Pageviews	46
5.9	Gipfelblog Outlook.....	48
6	Continent of Edu-Blogs Use Case: InternetWorking Blog and openHPI	49
6.1	Massive Open Online Courses.....	49
6.2	Supporting Cooperative Social Learning.....	51
6.2.1	Connecting Learners of MOOCs	52
6.2.2	Social Interaction in MOOCs	52
6.3	Use Case: InternetWorking-Blog	53
7	Continent of Ego-Blogs: Use Case – svenblogt.de.....	57
7.1	Interests	57
7.2	Writing Style	59
7.3	Audience.....	59
7.4	Activity	59
8	Continent of Corporate-Blogs: Use Case – SAP Blog	61
8.1	The Corporate Internal Communications Perspective	61
8.2	Deployment of Corporate Weblogs.....	62
8.3	Success Factors	63
8.4	Point of View (POV) Platform	64
8.4.1	POV: Scope and Motivation.....	65
8.4.2	Configuration of the Standardized to Fit Corporate Requirements.....	65
8.4.3	Who Are You Really?	67

8.4.4	Seamless Integration.....	68
8.4.5	Meeting Enterprise Standards.....	69
8.5	Proof of Concept and Outlook.....	70

Part III The Explorer’s Path Through the Blogosphere

9	The Challenge of Exploring the Blogosphere.....	75
9.1	Crawling.....	76
9.2	Analytics.....	77
9.2.1	New Application Areas.....	77
10	Towards an Exploration Machine for the Blogosphere.....	79
11	Data Extraction.....	83
11.1	Existing Approaches.....	84
11.2	Information Elements of Interest.....	84
11.3	Implementation Details.....	86
11.4	Optimization.....	89
11.4.1	Identification of Blogrolls.....	89
11.4.2	Identification of Trackbacks.....	90
11.4.3	Reliability of Feedparsing.....	90
11.4.4	Language Detection.....	91
11.4.5	Postlinks.....	91
11.4.6	Prioritization.....	92
11.4.7	News Portals.....	93
11.4.8	Matching of Twitter Accounts.....	93
11.5	Crawler Performance Summary.....	95
12	Data Analysis.....	101
12.1	Possible Analyses.....	101
12.1.1	Network Analysis.....	102
12.1.2	Content Analysis.....	104
12.2	Clustering the Blogosphere.....	105
12.2.1	Step 1: Data Extraction.....	105
12.2.2	Step 2: Data Preparation.....	106
12.2.3	Step 3: Data Aggregation.....	107
12.2.4	Step 4: Data Classification.....	107
12.2.5	Step 5: Calculation of Coordinates.....	111
12.2.6	Step 6: Visualization.....	112
12.3	Ranking the Blogosphere.....	112
12.3.1	Related Work.....	113
12.3.2	Discussion of Ranking Criteria.....	116
12.3.3	Analysis of Existing Blog-Ranking Services.....	121
12.3.4	Ranking Metric.....	125
12.3.5	Implementation and Validation.....	131
12.3.6	Limitations.....	134
12.3.7	Conclusion.....	134

13 Data Visualization	135
13.1 Related Research	136
13.2 PostConnect	137
13.2.1 Implementation of PostConnect	137
13.2.2 Visualization of PostConnect	139
13.2.3 Final Remarks About PostConnect	143
13.3 BlogConnect	144
13.3.1 Visualization of BlogConnect	145
13.3.2 Implementation of BlogConnect	148
13.3.3 Final Remarks About BlogConnect	152
13.4 TrendViz	153
13.4.1 Main Interface	153
13.4.2 Posts View	156
13.4.3 History View	156
13.5 Information Spreading	157
 Part IV The Nautilus of the Blogosphere: BLOGINTELLIGENCE	
14 The BLOGINTELLIGENCE Portal	161
14.1 Search and Custom Ranking	161
14.2 Visual Exploration Tools	164
14.2.1 Relations	164
14.2.2 Trends	166
14.2.3 Emotions	167
15 Analyzing and Forecasting Trends	171
15.1 Related Work	172
15.2 Trend Detection Preparation	174
15.2.1 Data	174
15.2.2 Term Extraction	176
15.2.3 Time Window	177
15.2.4 Importance Index	178
15.2.5 Term Clustering	180
15.3 Trend Detection Algorithm	182
15.3.1 Linear Regression	183
15.3.2 Content Analysis	184
15.3.3 Tag Analysis	185
15.3.4 Link Analysis	187
15.3.5 Trend Detection	188
15.4 Trend Detection Evaluation	190
15.4.1 Dataset for Evaluation	190
15.4.2 Trend Detection Preparation	191
15.4.3 Trend Detection Algorithm	195

15.4.4	Trend Prediction	204
15.4.5	Performance.....	205
15.5	Trend Detection Vision	206
15.5.1	Phrase Extraction	206
15.5.2	Sentiment Analysis	207
15.5.3	Performance.....	207
15.5.4	Time Shifting	208
15.5.5	User Input	208
15.6	Trend Detection Final Remarks	208
16	Judging Consistency and Expertise of Blogs	211
16.1	Related Work	213
16.1.1	General Rankings	213
16.1.2	Blog-Specific Rankings	214
16.1.3	Consistency-Related Rankings.....	214
16.2	Definition of the Topic Consistency Metric	216
16.2.1	Consistency Between Posts (Inter-post)	216
16.2.2	Internal Consistency of Posts (Intra-post)	219
16.2.3	Consistency Between Posts and Classification (Intra-blog).....	220
16.2.4	Consistency of Linking and Linked Blogs (Inter-blog) ...	221
16.2.5	Combined Topic Consistency Rank	222
16.3	Implementation of Topic Detection	223
16.3.1	Prerequisites	223
16.3.2	Clustering	225
16.4	Implementation of the Topic-Consistency Rank	225
16.4.1	Intra-post Consistency	226
16.4.2	Inter-post Consistency	226
16.4.3	Intra-blog Consistency	227
16.4.4	Inter-blog Consistency	227
16.4.5	BIIMPACT Score	228
16.5	Consistency Rank Evaluation	229
16.5.1	Experimental Setup.....	229
16.5.2	Clustering	229
16.5.3	Results of the Topic Consistency Sub Ranks.....	231
16.5.4	Comparison of BIIMPACT and Combined Topic Consistency Rank	233
16.6	Consistency Rank Future Research.....	236
16.6.1	Enhanced Topic Detection	236
16.6.2	Visualization	237
16.7	Consistency Rank Final Remarks	239
17	Vision of the Blogosphere and Its Exploration	241
17.1	Influence of the Blogosphere and Expected Growth.....	241
17.2	Data Extraction	242
17.3	Data Analysis.....	243

17.4	Visualization and Provision	244
17.4.1	PostConnect	245
17.4.2	BlogConnect	245
17.5	BLOGINTELLIGENCE Outlook	246
Bibliography		247
Index		267