

Elisabeth Günther

Topic Modeling

Algorithmische Themenkonzepte
in Gegenstand und Methodik der
Kommunikationswissenschaft

HERBERT VON HALEM VERLAG

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Elisabeth Günther

Topic Modeling.

Algorithmische Themenkonzepte in Gegenstand und Methodik der Kommunikationswissenschaft

Köln: Halem 2022

D6

ELISABETH GÜNTHER studierte Kommunikationswissenschaft in Augsburg (B.A.) und Hohenheim (M.Sc.). Von 2012-17 war sie wissenschaftliche Mitarbeiterin an den Universitäten Hohenheim und Münster. Ihre Forschungsinteressen liegen im Bereich Journalismusforschung und Computational Methods. Seit 2017 arbeitet sie als Data Scientistin.

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder unter Verwendung elektronischer Systeme (inkl. Online-Netzwerken) gespeichert, verarbeitet, vervielfältigt oder verbreitet werden.

© 2022 by Herbert von Halem Verlag, Köln

ISBN (Print): 978-3-86962-575-1

ISBN (PDF): 978-3-86962-569-0

Den Herbert von Halem Verlag erreichen Sie auch im Internet unter <http://www.halem-verlag.de>
E-Mail: info@halem-verlag.de

SATZ: Herbert von Halem Verlag

LEKTORAT: Rabea Wolf, Volker Manz

DRUCK: docupoint GmbH, Magdeburg

GESTALTUNG: Claudia Ott Grafischer Entwurf, Düsseldorf

Copyright Lexicon ©1992 by The Enschedé Font Foundry.

Lexicon® is a Registered Trademark of The Enschedé Font Foundry.

Inhalt

Abkürzungsverzeichnis	15
-----------------------	----

TEIL I RELEVANZ UND FRAGESTELLUNG

1.	Die duale Bedeutung von Topic Modeling für die KW	20
1.1	Forschungsleitendes Interesse	20
1.1.1	<i>Algorithmische Themen im Gegenstand der KW</i>	22
1.1.2	<i>Algorithmische Themen in der Methodik der KW</i>	23
1.1.3	<i>Methodologische Problemstellung: Vergleich der manuell-deduktiven und automatisch-induktiven Themenanalyse</i>	24
1.2	Zum Aufbau der Arbeit	26
1.3	Danksagungen	29

TEIL II ALGORITHMEN UND CODE

	Einleitung: Zur Bedeutung und Begründung der algorithmischen Logik	32
2.	Technologischer Wandel: Die Vierte Industrielle Revolution	33
2.1	Big Data	35
2.1.1	<i>Zwischen Gesellschaftswandel, Datenbankproblem und Pathos</i>	36
2.1.2	<i>Digitalisierung und Datafizierung</i>	39
2.1.3	<i>Algorithmische Logik als Treiber der Automatisierung</i>	43

2.2	Künstliche Intelligenz	47
2.2.1	<i>Wesen und Ziel der KI</i>	49
2.2.2	<i>Geschichte der KI</i>	56
2.2.3	<i>KI-Trends und Zukunftsprognosen</i>	60
2.3	Implikationen für Alltag und Gesellschaft	65
2.3.1	<i>Techno-Social Environment</i>	66
2.3.2	<i>Wie verändern Digitalisierung und Big Data die Choice Architecture?</i>	68
2.3.3	<i>Reverse-Turing-Test: Dehumanisierung im Techno-Social Environment</i>	72
2.3.4	<i>Ethik für Algorithmen</i>	78
2.4	Implikationen für die empirische Forschung	87
2.4.1	<i>Big Data und Algorithmen in der empirischen Forschung</i>	88
2.4.2	<i>Data Science und Computational Social Sciences</i>	94
2.4.3	<i>Kritik an Big Data und Co</i>	100
2.5	Zwischenfazit: Algorithmen und Code in der Makroperspektive	104
2.5.1	<i>Big Data und KI in Gegenstand und Methodik der KW</i>	104
2.5.2	<i>Zentrale Folgerungen für die Computational Social Sciences</i>	106
3.	Algorithmische Verarbeitung natürlicher Sprache	111
3.1	Sprache als Zeichensystem (de Saussure)	114
3.1.1	<i>Zentrale Relevanz der Sprache</i>	114
3.1.2	<i>Dyadischer Zeichenbegriff und Weiterführung</i>	116
3.2	Zeichenlehre als Erkenntnistheorie (Peirce)	118
3.2.1	<i>Zeichenbegriff und Grundlagen</i>	118
3.2.2	<i>Syntaktik, Semantik und Pragmatik: Weiterführung nach Morris</i>	123
3.2.3	<i>Übergang in die Stochastik: Deduktion, Induktion und Abduktion</i>	124
3.3	Information und Kommunikation (Shannon/Weaver)	127
3.3.1	<i>Von der Zeichenlehre zur Informationstheorie</i>	129
3.3.2	<i>Definition von Information als stochastischer Prozess</i>	133
3.4	Algorithmische Verarbeitung von Sprache und Information	137
3.4.1	<i>Natural Language Processing</i>	138
3.4.2	<i>Machine Learning</i>	142

3.5	Kognition und sprachliche Repräsentation	146
3.5.1	<i>Einführung in die Kognitionswissenschaft</i>	148
3.5.2	<i>Computational Theory of Mind</i>	151
3.6	Zwischenfazit: Algorithmen und Code in der Mikroperspektive	155
3.6.1	<i>Computationale Verarbeitung natürlicher Sprache</i>	155
3.6.2	<i>Mensch vs. Maschine?</i>	157
4.	Fazit: Die Omnipräsenz der algorithmischen Logik	160
4.1	Veränderte Bedingungen in Gegenstand und Methodik der Kommunikationswissenschaft	160
4.2	Leitmotive	162
4.2.1	<i>Motiv 1: Eine strikte Dichotomie aus ›Mensch vs. Maschine‹ greift fehl</i>	163
4.2.2	<i>Motiv 2: KI schafft keine ›Magic Buttons‹ – vom Ob zum Wie</i>	164
4.2.3	<i>Motiv 3: Die KW hat die Kompetenz und Verantwortung, im Diskurs um algorithmische Systeme eine zentrale Rolle einzunehmen</i>	166

TEIL III

DAS KONSTRUKT ›THEMA‹

Einleitung:

	Zwei Perspektiven auf das Konstrukt ›Thema‹	169
--	---	-----

5.	Das Thema als Einheit der öffentlichen Meinung	171
5.1	Was ist ›das Thema‹?	172
5.1.1	<i>Themenbegriff, -konstrukt, -konzept und -variable</i>	172
5.1.2	<i>Das Konstrukt ›Thema‹ in der KW-Forschungstradition</i>	173
5.2	Themen konstituieren Öffentlichkeit	175
5.2.1	<i>Drei Ebenen von Thematisierung</i>	175
5.2.2	<i>Integrationsfunktion von Themen</i>	177
5.3	Themenauswahl konstruiert Medienrealität	179
5.3.1	<i>Gatekeeper</i>	180
5.3.2	<i>News-Bias</i>	181
5.3.3	<i>Nachrichtenwert</i>	182

5.4	Einfluss und Beeinflussung massenmedialer Thematisierung	184
5.4.1	<i>Praeludium: Der Issue-Begriff in der politischen Kommunikation</i>	185
5.4.2	<i>Wirkungshypothese im Agenda-Setting-Ansatz</i>	187
5.4.3	<i>Issues als Gegenstand der strategischen Kommunikation</i>	189
5.4.4	<i>Neue Aspekte algorithmisch strukturierter Öffentlichkeit</i>	192
5.5	Thematische Ausdifferenzierung führt zu Fragmentierung?	195
5.5.1	<i>Ursprünge der Fragmentierungsthese</i>	196
5.5.2	<i>Stratifizierte und segmentierte Öffentlichkeit</i>	197
5.5.3	<i>Themen in der Filterblase</i>	199
5.5.4	<i>Kritik an der Fragmentierungsthese</i>	201
5.6	Zwischenfazit: Das Konstrukt ›Thema‹ in der Makroperspektive	205
5.6.1	<i>Die kommunikationswissenschaftliche Forschungstradition</i>	205
5.6.2	<i>Thematisierung in algorithmisch strukturierter Öffentlichkeit</i>	207
6.	Das Thema als Wissensstruktur	210
6.1	Was ist ›das Thema‹?	213
6.2	Textverstehen bottom-up: Die Textoberfläche	216
6.2.1	<i>Propositionsmodelle</i>	216
6.2.2	<i>Makrostrukturmodelle: Einzug des Themenbegriffs</i>	218
6.3	Textverstehen top-down: Schematheoretische Ansätze	221
6.3.1	<i>Schema (Bartlett)</i>	221
6.3.2	<i>Skripte (Schank/Abelson)</i>	223
6.3.3	<i>Frames (Minsky)</i>	224
6.3.4	<i>Konstruktions-Integrations-Modell (Kintsch)</i>	225
6.3.5	<i>Das Themenkonstrukt in den schematheoretischen Ansätzen</i>	226
6.3.6	<i>Interdisziplinäre Anleihen</i>	229
6.4	Text im Kontext: Textverstehen anhand mentaler Modelle	232
6.4.1	<i>Mentale Modelle (Johnson-Laird)</i>	233
6.4.2	<i>Situationsmodelle (van Dijk/Kintsch)</i>	235
6.4.3	<i>Das Themenkonstrukt in mentalen Modellen</i>	237

6.5	Zwischenfazit: Das Konstrukt ›Thema‹ in der Mikroperspektive	238
6.5.1	<i>Die Entwicklung des Forschungsbereichs Textverstehen</i>	238
6.5.2	<i>Themendefinition im Bereich des Textverstehens</i>	240
6.5.3	<i>Verbindung zum Forschungsbereich KI</i>	241
7.	Fazit:	
	Es gibt kein universales Verständnis von ›Thema‹	243
7.1	Unterschiedliche Ausgangspositionen für kw und KI	243
7.2	Zentrale Erkenntnisse zum Themenkonstrukt	244
7.2.1	<i>Themendefinitionen, Themenkonzepte</i>	244
7.2.2	<i>Ineinandergreifen der beiden Perspektiven</i>	246

TEIL IV ALGORITHMISCHE THEMEN

	Einleitung: Manuelle und algorithmische Themenanalysen	249
8.	Das Thema als Variable in der kommunikationswissenschaftlichen Forschungstradition	254
8.1	Das klassische Vorgehen	256
8.2	Vom Themenkonzept zur Themenvariable	257
8.2.1	<i>Kategorienbildung</i>	257
8.2.2	<i>Ereignisbezug von Themen</i>	259
8.3	Codierphase als gelenkte Rezeption	261
8.4	Computationale Verfahren im klassischen Forschungsprozess	264
8.5	Zwischenfazit: Die klassische kw-Forschungspraxis	265
9.	Algorithmische Themenanalyse	267
9.1	Computationale Verarbeitung natürlichsprachiger Texte	269
9.2	Latent Semantic Analysis (LSA)	270
9.2.1	<i>LSA als ›Theory of Meaning‹</i>	272
9.2.2	<i>LSA als Textanalyseverfahren</i>	273
9.2.3	<i>Mathematische Grundlagen</i>	274

9.3	Probabilistic Latent Semantic Analysis (PLSA)	281
9.3.1	<i>Generatives Modell</i>	282
9.3.2	<i>Schätzen der Modellparameter</i>	284
9.3.3	<i>Zentrale Errungenschaften im Topic Modeling</i>	287
9.4	Latent Dirichlet Allocation (LDA)	288
9.4.1	<i>LDA = PLSA + Bayes</i>	290
9.4.2	<i>Generatives Modell</i>	291
9.4.3	<i>Vorbereitung der Datenanalyse</i>	293
9.4.4	<i>Schätzen der Modellparameter</i>	294
9.5	Zwischenfazit: Die computationale Forschungspraxis	298
10.	Fazit: Ganzheitliche Gegenüberstellung der Themenanalysen	301
10.1	Mensch vs. Maschine	302
10.2	Zentrale Unterschiede in der Untersuchungsanlage	305
10.3	Konsequenzen für den Forschungsprozess	308
10.4	Konsequenzen für die Qualität der Messung	310
10.4.1	<i>Reliabilität als triadische Beziehung</i>	311
10.4.2	<i>Apparat</i>	313
10.4.3	<i>Kontext</i>	315
10.4.4	<i>Material</i>	317
10.5	Komplementäre Instrumente zur Themenanalyse	319

TEIL V

FAZIT UND DISKUSSION

11.	Topic Modeling ist fester Bestandteil der kw	321
11.1	Der Blick zurück	321
11.2	Integration der algorithmischen Themenanalyse in die kw	323
11.2.1	<i>These 1: Innovation und Allgegenwärtigkeit</i>	323
11.2.2	<i>These 2: Duale Relevanz</i>	324
11.2.3	<i>These 3: TM ≠ manuelle Themenanalyse + Automatisierung</i>	324
11.2.4	<i>These 4: Umfassende Integration</i>	326
11.2.5	<i>These 5: Menschliche Intelligenz</i>	327

11.2.6	<i>These 6: Konzeptionelle Skills</i>	327
11.2.7	<i>These 7: Kompetenz und Verantwortung</i>	329
11.3	Kritische Reflexion	330
11.4	Der Blick nach vorn	331
11.4.1	<i>Sapir-Whorf-Hypothese: Zum Einfluss der (künstlichen) Sprache</i>	332
11.4.2	<i>Computational Turn, Computational Divides?</i>	334
11.4.3	<i>Das menschliche Themenverständnis im Techno-Social Environment</i>	336
	Literaturverzeichnis	339