

Maximilian Wellmann

**Kommerzielles Text- und
Data-Mining im Urheberrecht**



PETER LANG

Gliederung

Vorwort	5
Gliederung	7
Abkürzungsverzeichnis	21
A. Einführung und Hintergrund	23
I. Problemaufriss – von der Informationsgesellschaft in die „data-mining society“	24
II. Anstoß und Gang der Arbeit	26
III. Forschungsfragen	27
IV. Technischer Hintergrund und begriffliche Grundlagen des Text- und Data-Mining	29
V. Begriffsbestimmungen und technischer Hintergrund der Informatik	29
1. Data-Mining	30
a) Alternative 1: Data-Mining als Methode – Knowledge Discovery in Databases als Prozess	31
b) Alternative 2: Data-Mining als Prozess und Methode	32
2. Text-Mining	32
3. Web-Mining	33
4. Text- und Data-Mining-Prozess	34
a) Analyseziel: Informationsgewinnung	35
b) Analyserohstoff: Daten	35
aa) Big Data	37
bb) Big Data und die Notwendigkeit automatisierter Datenanalyseverfahren	37
c) Zentrale Analysemethoden	38
aa) Klassifikation	39

bb) Segmentierung	39
cc) Abhängigkeitsanalyse	40
dd) Abweichungsanalyse	40
5. Einsatzgebiete des Text- und Data-Mining	41
a) Text- und Data-Mining in der Wissenschaft	42
b) Text- und Data-Mining im kommerziellen Sektor	42
c) Text- und Data-Mining im hoheitlichen Einsatz	43
VI. Begriffsbestimmung und technisches Verständnis der DSM-Richtlinie	44
1. Systematik	45
2. Legaldefinition des Art. 2 Nr. 2 DSM-RL	46
a) Schlagwortartiges Verständnis der DSM-Richtlinie	46
b) Prozessorientiertes Verständnis der DSM-Richtlinie	47
c) Reichweite des Datenbegriffs in Art. 2 Nr. 2 DSM-RL	48
d) Systematische Inkohärenz zu Art. 3, 4 DSM-RL	49
3. Kritik	50
VII. Begriffsbestimmung und technisches Verständnis des Urheberrechtsgesetzes	50
1. Systematik	50
2. Legaldefinition des § 44b Abs. 1 UrhG	51
3. Kritik	52
VIII. Terminologische Prämissen für den weiteren Fortgang der Arbeit	53

B. Die verwertungsrechtliche Analyse des zentralen Prozessschritts der Datenauswahl im Text- und Data-Mining	55
I. Die Untersuchung leitende Grundsätze der urheberrechtlichen Dogmatik	56
II. Datenauswahl aus Werken	57
1. Der einheitliche europäische Werkbegriff	57
a) Voraussetzungen	59

b) Der Schutz von Werkteilen	61
2. Eingriff in das Vervielfältigungsrecht durch die Auswahl von Werken	62
a) Gründe für das Eingreifen des § 16 UrhG	63
b) Grenzfälle	64
aa) Keine verbindliche Schutzuntergrenze durch die Rechtsprechung	65
bb) Konsequenz für das Text-Mining	66
c) Rechtliche Lösung: Die Datenauswahl als vorüberge- hende Vervielfältigung im Arbeitsspeicher	67
d) Technische Lösung: Die Methode der abgeleiteten Textformate	68
aa) Technischer Vorgang	68
bb) Kritik	69
III. Datenauswahl aus Datenbankwerken, § 4 Abs. 2 S. 1 UrhG	70
1. Besonderer Stellenwert der Datenauswahl aus Datenbanken für den Data-Mining-Prozess	71
2. Allgemeine Voraussetzungen des Datenbankbegriffs	72
a) Einordnung relationaler Datenbanken unter den Datenbankbegriff	76
aa) Technischer Aufbau	76
bb) Voraussetzungen des Datenbankbegriffs	78
(1) Problematische Voraussetzung der Unabhängigkeit	78
(2) Datenkombinationen als unabhängige Elemente nach EuGH „Esterbauer“	79
b) Einordnung eines Data Warehouse unter den Datenbankbegriff	80
aa) Technischer Aufbau	81
bb) Voraussetzungen des Datenbankbegriffs	81
c) Einordnung horizontal skalierten Datenbanksysteme unter den Datenbankbegriff	82
3. Besondere Voraussetzung: Persönlich geistige Schöpfung	83

4.	Eingriff in das Vervielfältigungsrecht durch den Vorgang der Datenauswahl	84
	a) Technischer Vorgang des Abrufs	84
	b) Abruf einzelner Daten	85
	c) Notwendige Differenzierung beim Abruf der gesamten Datenbank	85
IV.	Datenauswahl aus Datenbanken sui generis, §§ 87a ff. UrhG	86
1.	Besondere Voraussetzung: Quantitativ oder qualitativ wesentliche Investition	87
	a) Wesentliche Investition	87
	b) „Allerweltsinvestition“ oder signifikante Investition	88
2.	Eingriff in das Vervielfältigungsrecht durch die Datenauswahl	90
	a) Problemaufriss und Unterschiede in der Bewertung zum Abruf aus Werken und Datenbankwerken	92
	b) Abruf einzelner Daten	93
	c) Wiederholter und systematischer Abruf einzelner Daten	94
	aa) Die Datenauswahl als normale Nutzung, § 87b Abs. 1 S. 2 Alt. 1 UrhG	94
	bb) Keine unzumutbare Beeinträchtigung berechtigter Interessen, § 87b Abs. 1 S. 2 Alt. 2 UrhG	96
	d) Abruf des gesamten Datenbestandes aus einer Datenbank	97
3.	Sonderfall des in-database data mining	97
V.	Datenauswahl aus vertraglich geschützten Datenbanken	98
1.	EuGH „Ryanair“	99
2.	Rechtliche Implikationen der Datenauswahl	99
VI.	Der urheberrechtliche Schutz von Webseiten und seine Konsequenzen für die Datenauswahl	101
1.	Technische Struktur einer HTML-Webseite	101
2.	Urheberrechtlicher Schutz von Webseiten	102
	a) Webseiten als (Multimedia-)Werk	102
	b) Webseiten als Datenbankwerk	103
	c) Webseiten als Datenbank sui generis	103

aa) Einordnung für einzelne Benutzeroberflächen	104
bb) Einordnung für mehrere HTML-Webseiten und Plattformstrukturen	104
cc) Besondere Voraussetzung: Quantitativ oder qualitativ wesentliche Investition	105
d) Webseiten als Computerprogramm	106
3. Eingriff in das Vervielfältigungsrecht durch die Datenauswahl	106
a) Technischer Vorgang der Datenauswahl durch Screen- Scraping-Software	107
b) Abruf von Daten aus Webseiten	108
c) Abruf von Daten aus Webseiten als Datenbank sui generis	109
aa) Keine andere Wertung durch EuGH „Innoweb“	110
bb) Unwirksamkeit entgegenstehender vertraglicher Beschränkungen gem. § 87e UrhG	110
VII. Zwischenergebnis	111

C. Die verwertungsrechtliche Analyse der weiteren

Prozessschritte des Text- und Data-Mining	113
I. Datenvorbereitung (Preprocessing)	113
1. Preprocessing von Werken	114
a) (Weiterer) Eingriff in das Vervielfältigungsrecht	114
b) Kein Eingriff in das Bearbeitungsrecht durch Annotation und Reduktion des Datenbestandes	116
2. Preprocessing von Datenbanken	118
II. Transformation in eine Datenmatrix oder ein Textkorpus	118
1. Technische Vorgaben	119
2. Eingriff in das Vervielfältigungsrecht durch Konvertierung in ein maschinenlesbares Datenformat	120
a) Die urheberrechtliche Schutzfähigkeit einer Datenmatrix	121
aa) Schutzfähigkeit gem. §§ 2 Nr. 7, 4 Abs. 2 S. 1 UrhG	122

bb) Schutzfähigkeit als neue Datenbank gem. § 87a Abs. 1 S. 2 UrhG	122
(1) Ausschluss von Spin-off-Datenbanken durch die Rechtsprechung	124
(2) Übertragung auf die vorliegende Konstellation ...	125
b) Die urheberrechtliche Schutzfähigkeit eines Textkorpus	126
3. Eingriff in das Recht zur öffentlichen Wiedergabe bei Veröffentlichung der Datenmatrix	126
a) Konstellation 1	127
b) Konstellation 2	128
4. Eingriff in das Recht zur öffentlichen Wiedergabe bei Veröffentlichung des Textkorpus	129
III. Data-Mining im engeren Sinne: Die Anwendung des Data-Mining-Algorithmus	129
1. Technische Betrachtung	130
a) Konfigurierte Algorithmen	130
b) Selbstlernende Algorithmen	131
2. Urheberrechtliche Bewertung des Analysevorgangs	133
3. Urheberrechtliches Schicksal des Data-Mining-Algorithmus ..	133
a) Konfigurierte Algorithmen	133
b) Selbstlernende Algorithmen	134
IV. Evaluation und Interpretation der offenbaren Muster	135
V. Gesamtergebnis der verwertungsrechtlichen Analyse des Text- und Data-Mining-Prozesses	136
D. Die urheberrechtliche Schutzfähigkeit der Text- und Data-Mining-Ergebnisse	139
I. Abgrenzung zwischen Text- und Data-Mining-Ergebnissen und KI-Erzeugnissen	139
II. Schutzfähigkeit unter dem einheitlichen europäischen Werkbegriff	140

1. Der Anthropozentrismus des Werkbegriffs als Ausschlusskriterium	141
2. Möglichkeit eines derivativen Schutzes?	143
III. Schutzfähigkeit unter dem Schutzrecht sui generis des Datenbankherstellers	143
1. Voraussetzungen des § 87a Abs. 1 S. 1 UrhG	144
2. Berücksichtigungsfähigkeit der Investition	145
a) Abgrenzungspostulat nach EuGH „BHB/William Hill“	146
b) Problem der „selbstständigen Investition“ in mehrstufigen Datenanalyseprozessen	148
3. Teleologische Extension der Voraussetzung der Beschaffung in § 87a Abs. 1 S. 1 UrhG	150
a) Kritische Würdigung des Begriffs der Rohdaten	151
b) Ambivalente Auswirkung einer teleologisch-extensiven Auslegung auf das Text- und Data-Mining	152
c) Weitere Bedenken	153
4. Das Zusammenspiel des relativen und absoluten Datenbankschutzes und die Konsequenz der faktischen Exklusivität	154
IV. Der Schutz von Text- und Data-Mining-Ergebnissen im US-amerikanischen Copyright Law	155
1. Zielsetzung des Rechtsvergleichs	155
2. Schutzfähigkeit von Daten und Datensammlungen	156
3. Für das Text- und Data-Mining entscheidende Voraussetzungen des Werkschutzes	158
4. Misappropriation-Doktrin	159
a) Voraussetzungen	159
b) Anwendung auf die Ergebnisse des Text- und Data- Mining-Prozesses	161
5. Schutz in den USA erzeugter Text- und Data-Mining- Ergebnisse über §§ 87a ff. UrhG	162
a) Ansatz über § 127a Abs. 2 UrhG	162
b) Schutz durch Unterlizenzierung an europäische Tochtergesellschaften	163

V. Zwischenergebnis	164
E. Einordnung der Schrankenbestimmungen zum Text- und Data-Mining in den Acquis Unionaire	167
I. Entstehungsgeschichte	167
II. Die DSM-Richtlinie im Acquis Unionaire des Urheberrechts	169
1. Der Acquis Unionaire des Urheberrechts	170
2. Integration der DSM-Richtlinie	171
III. Defizite der bisher einschlägigen Rechtslage für das Text- und Data-Mining	172
1. Art. 5 Abs. 1 InfoSoc-RL	173
2. Art. 5 Abs. 3 lit. a InfoSoc-RL	173
3. Art. 6 Abs. 1, 2 lit. b Datenbank-RL	174
4. Bewertung des bisherigen Schrankenregimes	175
IV. Verhältnis der DSM-Richtlinie zur InfoSoc-Richtlinie	175
1. Art. 3, 4 DSM-RL als <i>leges speciales</i> zu den Schranken der InfoSoc- und Datenbank-Richtlinie	176
2. Gegenansicht	177
3. Bewertung	177
F. Materiellrechtliche Würdigung der Schrankenbestimmung des Art. 4 DSM-RL	179
I. Voraussetzungen, Art. 4 Abs. 1 DSM-RL	179
1. Persönlicher Anwendungsbereich	180
2. Sachlicher Anwendungsbereich	180
a) Keine Notwendigkeit einer extensiven Auslegung	181
b) Digitalisierung	182
c) Einzelnes Werk	183
3. Rechtmäßiger Zugang	183
a) Veröffentlichung von Schutzgegenständen im Internet ohne Zustimmung des Rechtsinhabers	184

b) Rechtfertigung der Voraussetzung über den Dreistufentest	186
4. Urheberpersönlichkeitsrechtliche Aspekte	187
II. Speicherung, Art. 4 Abs. 2 DSM-RL	187
III. Nutzungsvorbehalt, Art. 4 Abs. 3 DSM-RL	189
1. Sinn und Zweck des maschinenlesbaren Nutzungsvorbehalts	189
2. Problemaufriss	190
3. Robots.txt als maschinenlesbares Mittel i. S. v. Art. 4 Abs. 3 DSM-RL	191
a) Einordnung als technische Schutzmaßnahme	192
b) Einordnung als Erklärungszeichen	193
4. Rechtsfolgen der Erklärung eines maschinenlesbaren Nutzungsvorbehalts	194
5. Probleme im Zusammenhang mit Art. 7 Abs. 2 S. 2 DSM-RL	194
IV. Flankierender Rechtsvorbehalt, Umkehrschluss aus Art. 7 Abs. 1 DSM-RL	195
V. Auswirkungen des Private Ordering	196
1. Drohender überschießender Schutz	197
2. Auswirkung des Private Ordering auf die angeordnete Vergütungsfreiheit	198
3. Abgleich mit empirischen Untersuchungen	200
4. Lösung des Problems über ein bereichsspezifisches Zugangsrecht zu analyserelevanten Daten?	201
a) Konstruktion über das bestehende Datenbankherstellerrecht	202
b) Dogmatische Bedenken	202
VI. Public-Private Partnership: Verhältnis zu Art. 3 DSM-RL	204
VII. Vorläufige Bewertung	205

G. Die Umsetzung der Schranke zum kommerziellen Text- und Data-Mining in das Urheberrechtsgesetz	207
I. Umsetzungsspielraum durch den Harmonisierungsgrad	207
II. Ausnahme oder Beschränkung	209
III. Umsetzung in § 44b Abs. 2, 3 UrhG	210
1. Systematik	210
2. Voraussetzungen, § 44b Abs. 2 S. 1 UrhG	211
a) Vervielfältigungen von Werken	211
aa) Keine überschießende Umsetzung durch § 44b Abs. 2 S. 1 UrhG	212
bb) Richtlinienkonforme Auslegung im Hinblick auf Art. 15 Abs. 1 DSM-RL	213
cc) Keine richtlinienkonforme Auslegung im Hinblick auf weitere Verwertungsrechte	214
b) Rechtmäßiger Zugang	214
3. Löschpflicht, § 44b Abs. 2 S. 2 UrhG	215
a) Auslegungsprämissen	215
aa) Zeitpunkt	216
bb) Umfang	217
cc) Rechtsfolgen eines Verstoßes gegen die Löschpflicht	217
b) Konzeptionelle Schwäche der Löschpflicht	218
4. Nutzungsvorbehalt, § 44b Abs. 3 S. 1 UrhG	219
a) Begriff des Nutzungsvorbehalts	220
b) Problematische Aspekte	220
c) Unwirksamkeit des Nutzungsvorbehalts	221
5. Maschinenlesbarer Nutzungsvorbehalt, § 44b Abs. 3 S. 2 UrhG	222
a) Diskriminierungsverbot	223
b) Problematik der fehlenden Identität zwischen Rechtsinhaber und Webseitenbetreiber	223
aa) Praxis der Rechtseinräumung im Internet	224

bb) Zurechnung des erklärten Nutzungsvorbehalts über konkludente Vollmachtserteilung	225
cc) Voraussetzungen und Grenzen	225
dd) Übertragung auf § 44b Abs. 3 S. 2 UrhG	227
(1) Voraussetzungen	227
(2) Bewertung	228
c) Fehlende Techniksensibilität des § 44b Abs. 3 S. 2 UrhG	229
d) Unwirksamkeit maschinenlesbarer Nutzungsvorbehalte	230
6. Vergütungsproblematik	231
IV. Verhältnis zu weiteren Schrankenregelungen des UrhG	232
1. § 44a UrhG	232
2. § 87c Abs. 1 S. 1 Nr. 4 UrhG	232
3. § 69d Abs. 4 UrhG	233
V. Anpassung des § 95b Abs. 1 S. 1 Nr. 1, Abs. 3 Nr. 1 UrhG	234
VI. Konzeptionelle Schwäche der Schrankenregelung am Beispiel der Presse	235
1. Grundsätze der Rechtsprechung des EuGH	235
2. Keine Übertragung der Grundsätze auf Art. 4 DSM-RL	236
3. Kein Ausweichen auf §§ 50, 63 Abs. 1, 2 S. 1 UrhG	237
4. Kritik	238
VII. Bewertung des § 44b UrhG unter Berücksichtigung der Erkenntnisse zu Art. 4 DSM-RL	239
1. Stärkung der Rechtsinhaber durch die drei Instrumente des Private Ordering	239
2. Keine Beseitigung der rechtlichen Fragmentierung durch die DSM-Richtlinie	240
3. Weitere identifizierte Schwächen der Schranke	241
a) Regelungslücke im Hinblick auf die fehlende Identität zwischen Rechtsinhaber und Webseitenbetreiber	241
b) Fehlende Techniksensibilität	241
c) Anreizfunktion zur Abschottung gemeinfreier Inhalte	242
4. Die Schranke als Ausdruck einer problematischen verwertungsrechtlichen Einschätzung des Gesetzgebers	243

5. Gesamtfazit	244
----------------------	-----

H. Alternative Überlegungen zur Einordnung des Text- und Data-Mining	247
I. Alternative auf Schrankenebene: Die Fair-Use-Doktrin des US-amerikanischen Copyright Law	247
1. Die Voraussetzungen der Fair-Use-Doktrin und ihre Anwendung auf das Text- und Data-Mining	248
a) Erster Faktor	250
b) Zweiter Faktor	253
c) Dritter Faktor	254
d) Vierter Faktor	255
2. Vergleich der Fair-Use-Lösung zu Art. 4 DSM-RL	257
a) Keine allgemeine Unterlegenheit eines enumerativen Regelungsansatzes	257
b) Identifizierte Schwächen des Art. 4 DSM-RL als „Wettbewerbsnachteil“ zu Fair-Use-Regimen	258
II. Alternativen auf Schutzbereichsebene	259
1. Text- und Data-Mining als freier digitaler Werkgenuss	260
a) Idee des freien digitalen Werkgenusses	260
b) Verwertungsrechtliche Implikationen	262
2. Freistellung des Text- und Data-Mining über § 5 Abs. 1 UrhG, §§ 24 UrhG a. F., 23 Abs. 1 UrhG	263
a) Dogmatische Anknüpfung an § 5 Abs. 1 UrhG	263
b) Dogmatische Anknüpfung an §§ 24 UrhG a. F., 23 Abs. 1 UrhG	264
c) Bewertung	266
3. Idee eines normativen Vervielfältigungsrechts zur Freistellung des Text- und Data-Mining auf Schutzbereichsebene	266
a) Konstruktion eines normativen Vervielfältigungsrechts des Art. 2 lit. c InfoSoc-RL in EuGH „Metall auf Metall“	267
b) Übertragbarkeit auf die sonstigen Vervielfältigungsrechte in Art. 2 InfoSoc-RL	269

aa) Pro	269
bb) Contra	269
c) Bewertung	270
I. Vorschlag eines neuen § 16 Abs. 3 UrhG n. F.	273
I. § 23 UrhG Abs. 3 S. 3 UrhG als regelungstechnisches Vorbild	274
II. Konkrete Ausgestaltung des § 16 Abs. 3 UrhG n. F.	274
III. Erläuterungen	275
1. § 16 Abs. 3 S. 1 UrhG n. F.	275
2. § 16 Abs. 3 S. 2 UrhG n. F.	275
3. § 16 Abs. 3 S. 3 UrhG n. F.	276
IV. Vorteile der Lösung gegenüber der eingeführten Schranke des § 44b UrhG	277
J. Ergebnisse und Schlussbetrachtung	279
I. Zusammenfassung der Ergebnisse in Thesen	279
II. Wertende Schlussbetrachtung	283
Literaturverzeichnis	285