

Text und Data Mining für Forschungszwecke und Bibliotheken – KI in und durch Bibliotheken

Elke Brehm

Einführung

Text und Data Mining (TDM) ist als Forschungsmethode etabliert und der Bedarf Verlagspublikationen für TDM zu nutzen ist hoch. Es stellt sich im Rahmen der Diskussion um TDM die Frage, in welchem Verhältnis TDM zu Künstlicher Intelligenz (KI) steht: Ist jede KI-Anwendung auch immer gleich eine TDM-Anwendung? TDM kann auch mit Unterstützung von KI durchgeführt werden, z.B. können Sprachmodelle trainiert werden um einen Datenkorpus zu analysieren und neue Muster, Trends und Korrelationen zu finden (Ziel von TDM). Das Training von Sprachmodellen kann aber auch der Entwicklung neuronaler Netze dienen, mit denen nicht zwingend nur ein Ziel verfolgt wird. Text und Data Mining ist daher nur ein Teilbereich von KI: Nicht alles was sich unter dem Begriff KI entwickelt und als KI eingeordnet wird, ist gleichzeitig TDM.¹ Maßgeblich sind der mit der Anwendung verfolgte Zweck und das konkret vorliegende Szenario.² Für diesen Artikel werden nur die Teilbereiche von KI und TDM betrachtet, die von Relevanz für die Anwendung der urheberrechtlichen Regelungen zum TDM sind. Im Folgenden wird in diesem Artikel zwar nur noch der Begriff TDM verwendet. Erfasst ist jedoch auch KI, sofern in diesem Zusammenhang relevant.

Die Verwendung von KI oder TDM birgt neben dem Bedarf der Forschenden auch für viele Arbeitsprozesse in Bibliotheken ein großes Potential. Mit TDM können zum Beispiel eigene Bestände für Provenienzforschung analysiert, die Katalogisierung und Sacherschließung automatisiert werden.

Bibliotheken können mit Hilfe von KI aber auch Sprachmodelle zur (Weiter-) Entwicklung eigener Dienstleistungen trainieren. So können z.B. Kunden außerhalb der Öffnungszeiten beraten und neue Arten der Wissensrepräsentation erarbeitet, KI-Anwendungen in Rechercheportale und Discoverysysteme implementiert werden, um die Rechercheergebnisse zu verbessern. Bibliotheken können neue Dienstleistungen entwickeln, die Forschenden und anderen Kunden TDM und/oder das Training von Sprachmodellen mit der in Bibliotheken vorhandenen Fachliteratur ermöglichen.

1 Brand, Jonathan: Copyright Implications of the Relationship Between Generative Artificial Intelligence and Text and Data Mining, BLOG infojustice, 27.10.2023, < [, Stand: 08.02.2024.](https://infojustice.org/archives/45509#:~:text=%E2%80%9CTDM%20is%20a%20crucial%20first,of%20data%20or%20experience%2C%20without%20)

2 ebd.

Abstract

Text und Data Mining (TDM) ist eine anerkannte wissenschaftliche Methode und zählt zur „Künstlichen Intelligenz“. Neben anderen Arten von (Forschungs-) Daten können insbesondere auch wissenschaftliche Publikationen wertvolle Bestandteile des Textkorpus sein, das für TDM oder als Trainingsmaterial für KI genutzt wird. Das TDM für Texte nimmt in der Diskussion um Künstliche Intelligenz deshalb einen so großen Raum ein, weil die Texte urheberrechtlich geschützt sind oder sein können. Die Publikationen stehen in gedruckter Form in den Lesesälen von Bibliotheken und werden durch Bibliotheken in Form von digitalen Datenbanken und Zeitschriftenpaketen lizenziert und Forschenden zur wissenschaftlichen Nutzung zur Verfügung gestellt.

Im Sommer 2021 wurden die Schrankenregelungen zum TDM im deutschen Urheberrecht im Hinblick auf die Anforderungen der DSM-Richtlinie der EU überarbeitet. Sie begünstigen nun auch explizit Bibliotheken. Aber was erlauben die neuen Schrankenregelungen konkret? Dürfen die analogen und digitalen Bestände der Bibliotheken auch für TDM genutzt werden? Wie steht es mit digitalen Ressourcen von Verlagen, insbesondere aus dem EU-Ausland? Welche Rolle kommt insbesondere wissenschaftlichen Bibliotheken in diesem Kontext zu? Im Arbeitspaket S-7-2 der Base Service Measure S7 von NFDI4Ing wurden „Guidelines zum Text und Data Mining für Forschungszwecke in Deutschland“ erarbeitet, in denen die neuen Schrankenregelungen im Hinblick auf wissenschaftliche Fachliteratur im Detail analysiert wurden. Die Ergebnisse werden hier präsentiert.

Text and Data Mining (TDM) is an established scientific method and forms part of "Artificial Intelligence". In addition to other types of (research) data, especially scientific publications may be an essential element of the text corpus, which may be used for TDM or as training material for AI. TDM takes up a lot of space in the discussions about AI, because the texts are or may be protected by copyright. The publications are accessible in printed form in the reading rooms of libraries and access to licensed digital databases and journal packages is provided by libraries for researchers for scientific use.

In the summer of 2021 the statutory exception for TDM in German copyright law was revised in accordance with the requirements of the DSM-Guidelines of the EU. Now libraries are explicitly named as beneficiaries of the regulation. What do the new and revised statutory exceptions allow exactly? May the print or digital resources of libraries be used for TDM? What about digital resources of publishers, especially if they are located outside EU? What is the role of scientific libraries in this context?

Within the work package S-7-2 of Base Service Measure S7 of NFDI4Ing, „Guidelines for Text and Data Mining for Research Purposes in Germany“ were developed, which contain a detailed analysis of the new and revised statutory exceptions for TDM in German copyright law with regard to scientific publications. The results are presented here.

Zum 7.6.2021 sind die Schrankenregelungen des deutschen Urheberrechts zum Text und Data Mining (TDM-Schranken) neu erlassen worden. Sie basieren auf der Umsetzung der Vorgaben der „Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG“³ (DSM-Richtlinie) der Europäischen Union. Sie sind durch das „Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes“ vom 31.5.2021⁴ in deutsches Recht umgesetzt worden. Mit dem Erlass neuer Regelungen ist Unsicherheit verbunden, da viele Einzelfragen zur Auslegung für die Anwendung in der Praxis noch nicht geklärt sind. Das gilt insbesondere, weil es sich um neue Technologien handelt. Das Projektconsortium NFDI4Ing, Base-Services S7, hat die Problematik in Arbeitspaket S-7-2 aufgegriffen und eine Analyse der rechtlichen Rahmenbedingungen für TDM für wissenschaftliche Fachliteratur erarbeiten lassen, die insbesondere Verlagspublikationen in den Fokus nimmt. Diese Analyse ist unter dem Titel „Guidelines zum Text und Data Mining für Forschungszwecke in Deutschland“⁵ erschienen. Für die Beurteilung der international privatrechtlichen Situation wurde ein Rechtsgutachten eingeholt, das ebenfalls veröffentlicht ist.⁶ Die Guidelines erläutern die neuen urheberrechtlichen Rahmenbedingungen für Text und Data Mining in Deutschland.

Dieser Aufsatz ist eine Zusammenfassung des bei der BiblioCon 2023 gehaltenen Vortrags mit dem Titel „Text und Data Mining für Forschungszwecke und Bibliotheken“⁷. Im Vortrag wurden die Ergebnisse der im Rahmen der Initiative NFDI4Ing erschienenen Guidelines aus der Perspektive der Bibliotheken dargestellt. Im Rahmen dieses Aufsatzes wird jeweils auf die vertiefte Darstellung in den Guidelines verwiesen, die umfangreiche Hinweise auf (Fach-) Literatur enthalten.

Die Darstellung beschränkt sich auf Szenarien, in denen wissenschaftliche Fachliteratur für TDM zu Forschungszwecken genutzt wird.

Um dem Bedarf der Forschenden zu begegnen, haben viele Verlage bereits vor Erlass der neuen TDM-Schranken

über ihre Webseiten Angebote zum TDM veröffentlicht, die zu den dort genannten Bedingungen genutzt werden können. Diese Angebote werden hier nicht weiter geprüft. Im Folgenden wird nach einer allgemeinen Einführung zum TDM der Anwendungsbereich der neuen TDM-Schranken, das Zusammenspiel zwischen TDM-Schranken, Verträgen und Handlungsort bei lizenzierten Inhalten und eventuell bestehende Risiken erläutert. Im Anschluss werden der Inhalt und die Voraussetzungen der TDM-Schranken dargestellt. Zum Abschluss wird der mögliche Handlungsrahmen für und die Rolle von Bibliotheken bei TDM für eigene Zwecke und für Forschungszwecke skizziert.

Allgemeines zum Text und Data Mining

Die im Folgenden dargestellten Regelungen zum TDM beruhen auf der DSM-Richtlinie der EU, die in den neuen Paragraphen §§ 44b und 60d neue Fassung UrhG⁸ in deutsches Recht umgesetzt wurde. Die Regelungen sind bis auf geringfügige, aber mitunter bedeutsame Abweichungen in allen EU-Mitgliedsstaaten gleich. § 44b Abs. 1 UrhG enthält die Legaldefinition für TDM: Es ist die „automatisierte Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen“.

Das Ziel von TDM, das durch die neuen Schrankenregelungen begünstigt sein soll, ist, neue und bisher verborgene Muster und Korrelationen zu finden.⁹ Dieses Ziel wird erreicht durch die automatisierte Analyse einer großen Anzahl an Texten bzw. Daten oder eines großen Datenkorpus, in dem Texte und Daten kombiniert sind.¹⁰

Wichtig ist für Bibliotheken, den Anwendungsbereich der TDM-Schranken von anderen speziell Bibliotheken begünstigenden Schrankenregelungen abzugrenzen: Zu nennen ist hier die von § 60e Abs. 1 UrhG begünstigte Vervielfältigung zur Indexierung z.B. zwecks Verbesserung der Volltextsuche, die aber eben nicht die Suche nach Mustern, Trends und Korrelationen zum Ziel hat und daher kein TDM im Sinne der TDM-Schranken darstellt.

3 Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, Amtsblatt der Europäischen Union L130/92, Online: <<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32019L0790&from=DE>>.

4 Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31.05.2021, Bundesgesetzblatt 2021 Teil I Nr. 27 S. 1204, Online: <http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl121s1204.pdf>.

5 Brehm, Elke: Guidelines zum Text und Data Mining für Forschungszwecke in Deutschland, 28.10.2022, Online: <<https://oa.tib.eu/renate/handle/123456789/10352>>.

6 Jani, Ole; Vonthien, Maximilian: Gutachten zu bestimmten kollisionsrechtlichen Fragen zur umfassenden Nutzung von lizenzierten Bibliotheksbeständen zum Zwecke des Text und Data Mining, 2022, Online: <<https://oa.tib.eu/renate/handle/123456789/10340>>.

7 Brehm, Elke: Text und Data Mining für Forschungszwecke und Bibliotheken, Vortrag bei der 111. BiblioCon2023, Online: <[urn:nbn:de:0290-opus4-184131](https://nbn:de:0290-opus4-184131)>.

8 Urheberrechtsgesetz vom 9. September 1965 (BGBl. I S. 1273), zuletzt geändert durch Artikel 25 des Gesetzes vom 23. Juni 2021 (BGBl. I S. 1858).

9 Dreier, Thomas; Schulze, Gernot (Hg.); Dreier in: Urheberrechtsgesetz, § 44b Rn. 5, Verlag C. H. Beck, 7. Auflage 2022.

10 Vertiefend: Brehm, Guidelines, S. 10 ff., 2022.

Die neuen Schranken zum TDM richten sich an zwei separate Zielgruppen:

Für die nichtkommerzielle wissenschaftliche Forschung ist TDM in § 60d UrhG neue Fassung geregelt, für alle anderen Nutzergruppen und Nutzungszwecke findet die allgemeine TDM-Schranke in § 44b UrhG Anwendung. Auch Forschende, die nichtkommerzielle Forschungszwecke verfolgen, dürfen sich bei Einhaltung der Voraussetzungen auf § 44b UrhG berufen.

Lizenzverträge und TDM

Die Bereitstellung von Fachliteratur in Datenbanken, E-Book- und Zeitschriftenpaketen erfolgt auf der Basis von Lizenzverträgen zwischen Rechtsinhaberinnen/-inhabern und Bibliotheken. Anhand der Verträge muss geprüft werden, ob die TDM-Schranken eingreifen. Verhältnismäßig einfach ist die Rechtslage, wenn in den Verträgen die Geltung deutschen Rechts vereinbart wurde.¹¹ Oft haben die Rechtsinhaberinnen und -inhaber ihren Sitz jedoch außerhalb der EU. Urheberrecht ist aber nationales Recht, das auch in der EU nur punktuell durch EU-Rechtsetzung europaweit vereinheitlicht wurde. Daher stellt sich die Frage,

ob und wie die neuen TDM-Schranken sich auf Regelungen in diesen Verträgen und die Zulässigkeit des TDM, das auf Basis der TDM-Schranken durchgeführt wird, auswirken.

In den Verträgen mit Rechtsinhaberinnen und -inhabern im EU-Ausland ist meist entweder

- keine Regelung zu Gerichtsstand und anwendbarem Recht, oder
- die Geltung ausländischen (außereuropäischen) Rechts und ein ausländischer (außereuropäischer) Gerichtsstand, oder
- das Recht am Sitz des Beklagten und der Beklagtengerichtsstand, oder
- eine Schiedsklausel vereinbart.¹²

Eine Regelung zum Gerichtsstand bedeutet einerseits, dass ein bestimmtes Gericht international zuständig ist, hat aber andererseits automatisch zur Konsequenz, dass dieses Gericht sein eigenes nationales Prozess- und Kollisionsrecht anwendet.¹³

Sofern vorhanden, bestimmen die Parteien mit der Auswahl eines bestimmten nationalen Rechts (Rechtswahl)

¹¹ Brehm, Guidelines, S. 70 ff., 2022.

¹² Brehm, Guidelines, S. 90 ff., 2022.

¹³ Brehm, Guidelines, S. 91 f., 2022.

EINFACH UNKOMPLIZIERT

Flexibler Zahlungskomfort für Ihre Kunden

Mit innovativen Kassensystemen und Zahlungssystemen von CSG Systems bieten Sie Ihren Kunden alle Freiheiten beim Bezahlen. Unsere Software mit vielfältigen Schnittstellenlösungen zu Fachverfahren arbeitet für Sie. Für Ihre Bibliothek haben wir die maßgeschneiderte Lösung. Mehr unter www.csg-systems.de.

Kompakt und leistungsstark:

VarioPay oder BGT – wir haben den perfekten Automaten für Ihre Bibliothek.

CSG
systems

BiblioCon 2024
Stand 37

smart ideas that work for you

über das anwendbare materielle Recht.¹⁴ Internationales Privatrecht (Kollisionsrecht) bestimmt unter anderem, ob eine Rechtswahl wirksam ist und welches Recht zur Anwendung kommt, wenn keine (wirksame) Regelung über das anwendbare Recht getroffen wurde. Das Kollisionsrecht ist nationales Recht und kann daher unterschiedlich ausgestaltet sein.¹⁵

Die Regelung über den Gerichtsstand und das anwendbare Recht können großen Einfluss auf den Ausgang eines Prozesses haben.

Bei internationalen Sachverhalten gilt im Internationalen Privatrecht das Schutzlandprinzip. Es bedeutet zusammengefasst, dass die Frage, ob eine Handlung ein Recht verletzt, immer nach dem Recht des Ortes beurteilt werden muss, an dem die Handlung vorgenommen wird. Daher sind nach dem Schutzlandprinzip die Schrankenregelungen des deutschen Urheberrechts unmittelbar anwendbar, wenn die maßgeblichen Handlungen in Deutschland stattgefunden haben.¹⁶ Das gilt auch für Verträge mit Rechtsinhaberinnen und -inhabern im EU-Ausland. Eine Vereinbarung von ausländischem Recht und die Bestimmung eines Gerichtsstands haben daher keinen Einfluss auf Zulässigkeit der Durchführung des TDM an sich, sofern die nachfolgenden Empfehlungen eingehalten werden.

Empfehlung – Durchführung von TDM in Deutschland

Grundsätzlich ist zu empfehlen, alle Nutzungshandlungen aller am TDM Beteiligten in Deutschland vornehmen zu lassen. Dann ist TDM nach folgender Maßgabe auf der Basis der deutschen TDM-Schranken zulässig:

Bei Vertragsschluss vor dem 1.3.2018 ist TDM zu wissenschaftlichen Zwecken nach der jeweils im Zeitpunkt der Vornahme des TDMs geltenden Schrankenregelung zulässig, sofern der Vertrag es nicht verbietet und TDM nach dem 1.3.2018 durchgeführt wird.¹⁷ TDM zu kommerziellen Zwecken ist nicht gestattet.¹⁸

Bei Vertragsschluss zwischen dem 1.3.2018 und 7.6.2021 ist TDM zu wissenschaftlichen Zwecken nach der jeweils im Zeitpunkt der Vornahme des TDMs geltenden Schrankenregelung zulässig. Das gilt selbst dann, wenn der Vertrag es verbietet. TDM zu kommerziellen Zwecken ist hingegen nicht gestattet.¹⁹

Bei Vertragsschluss nach dem 7.6.2021 ist TDM zu jedem (legalen) Zweck bei Einhaltung der Voraussetzungen der jeweils anwendbaren Schrankenregelungen zulässig.²⁰

Folgen der Vornahme einzelner Handlungen im Ausland

Sofern nicht sämtliche Nutzungshandlungen aller Beteiligten in Deutschland stattfinden, kann nach dem Schutzlandprinzip das Recht eines anderen Staates zur Anwendung kommen. Während innerhalb Europas die Rechtslage durch die zwingende Umsetzung der EU-Richtlinie in nationales Recht weitgehend vereinheitlicht wurde, gibt es in Ländern außerhalb Europas (bisher) nur vereinzelt Regelungen zu TDM. Diese begünstigen meist allein die Forschenden selbst, nicht aber Bibliotheken, Forschungs- oder Infrastruktureinrichtungen.

Es bleibt daher ein Risiko, auch nur einzelne Handlungen des TDM außerhalb von Deutschland durchführen zu lassen. Die Durchführung von TDM könnte nach dem jeweils anwendbaren nationalen Recht eine Urheberrechtsverletzung darstellen, selbst wenn der durch die im deutschen Urheberrecht verankerten TDM-Schranken gesetzte Rahmen eingehalten wird.²¹

Risiken bei Verträgen mit Rechtsinhaberinnen/-inhabern im EU-Ausland

Zwar ist, sofern sämtliche für TDM erforderliche Handlungen nach dem 1.3.2018 in Deutschland vorgenommen werden, TDM in jedem Fall zulässig und kann nicht verhindert werden. Bei Verträgen mit Rechtsinhaberinnen und -inhabern im EU-Ausland können jedoch Schadensersatzansprüche bestehen: Denn TDM kann zwar durch vertragliche Regelungen nicht verboten werden. Aber die Missachtung der vertraglichen Verbote kann eine Vertragsverletzung darstellen, die sich international privatrechtlich nach dem Vertragsstatut (dem im Vertrag vereinbarten oder durch internationales Privatrecht bestimmten nationalen Recht) richten würde.²² Denkbar ist, dass das nach dem dann anwendbaren nationalen Recht die Rechtssituation in Deutschland (und die Zulässigkeit des TDM in Deutschland) nicht beachtet werden muss. Das Schutzlandprinzip ist für die Frage, ob eine Vertragsverletzung vorliegt, nicht anwendbar.

¹⁴ Brehm, Guidelines, S. 93, 2022.

¹⁵ Brehm, Guidelines, S. 92 f., 2022.

¹⁶ Brehm, Guidelines, S. 93 f., 2022.

¹⁷ Brehm, Guidelines, S. 70 ff. und 74, 2022.

¹⁸ Zu typischen Klauseln in Verträgen: Brehm, Guidelines, S. 70 ff., 2022.

¹⁹ Brehm, Guidelines, S. 74, 2022. Zu § 60d UrhG in der zwischen dem 1.3.2018 und 7.6.2021 geltenden Fassung („alte Fassung“): Brehm, Guidelines, S. 88 ff., 2022.

²⁰ Brehm, Guidelines, S. 74, 2022.

²¹ Brehm, Guidelines, S. 96, 2022.

²² Brehm, Guidelines, S. 96 f., 2022.

Planung des TDM

Unter rechtlichen Aspekten ist folgendes Vorgehen zu empfehlen:

Vor Beginn der Durchführung von TDM müssen die Verträge geprüft werden: Der Vertragsschluss muss nach 1.3.2018 liegen, und darf kein explizites Verbot von TDM enthalten und auch keine Bedingungen, die als Verbot auszulegen sind.

Zulässig ist nur das, was die TDM-Schranken erlauben: Vom Umfang her sind die Vorgaben der TDM-Schranke zu berücksichtigen, die im Zeitpunkt der Vornahme der Nutzungshandlung gilt oder galt. Sofern keine TDM-Schranke bei Vornahme der Handlung galt, ist TDM nicht zulässig. Sofern Inhalte unter Open-Access-Lizenzen bereitgestellt sind, ist der Handlungsspielraum größer, weil die Open-Access-Lizenzen ggfls. mehr Handlungsspielraum erlauben, als die TDM-Schranken.²³

Urheberrechtsfreie Inhalte können frei genutzt werden.²⁴ Wenn TDM-Schranken nicht anwendbar sind, besteht die Möglichkeit, die Inhalte als abgeleitete Textformate im Rahmen der freien Benutzung nach § 23 Abs. 1 S. 2 UrhG für TDM zu nutzen.²⁵

Inhalt der Schrankenregelungen zum Text und Data Mining

TDM zu nichtkommerziellen Forschungszwecken (§ 60d UrhG neue Fassung)

Vertragliche Einschränkungen der TDM-Schranke sind unbeachtlich, sofern TDM zu nichtkommerziellen Forschungszwecken durchgeführt wird (§ 60g Abs. 1 UrhG). Irrelevant ist daher ein vertragliches Verbot, wenn TDM in Deutschland durchgeführt wird. Vertragliche Beschränkungen können zum Beispiel der Ausschluss bestimmter Publikationen von TDM, Festlegung besonderer Entgelte für TDM, Ausschluss von durch die TDM-Schranke begünstigten Einrichtungen oder Kunden, oder Ähnliches sein. Begünstigt sind Forschungsorganisationen und ihre Bibliotheken, Kulturerbeeinrichtungen wie z.B. öffentlich zugängliche Bibliotheken und Museen, Archive und Einrichtungen des Ton- und Filmerbes und neuerdings auch unabhängige Forschende, sofern sie nichtkommerzielle Forschungszwecke verfolgen.²⁶ Die begünstigten Einrichtungen und Personen dürfen auch Sprachmodelle für TDM trainieren, so lange dies für nichtkommerzielle For-

schungszwecke erfolgt und keine darüber hinaus gehenden Ziele verfolgt werden.

Vom Umfang her dürfen im Rahmen des TDM von rechtmäßig zugänglichem Ausgangsmaterial, das urheberrechtlich geschützt ist, Kopien angefertigt, für TDM aufbereitet und automatisierte Analysen durchgeführt werden. Von der TDM-Schranke werden dabei alle Werkarten und auch im Rahmen von verwandten Schutzrechten geschützte Werke erfasst, sofern ein Verweis auf § 60d UrhG im Gesetzestext vorhanden ist. Analog vorliegendes Material darf digitalisiert und technisch bedingte Änderungen am Ausgangsmaterial vorgenommen werden.²⁷

Eine Online-Bereitstellung des Textkorpus ist nur für einen bestimmt abgegrenzten Kreis von Personen für deren gemeinsame wissenschaftliche Forschung (mit Nutzerauthentifizierung) und einzelne Dritte zur Überprüfung der Qualität der wissenschaftlichen Forschung erlaubt. Dies ist auch nur für Bibliotheken und Forschende gestattet, die einer Forschungsorganisation angehören, nicht unabhängige Forschende. Eine Online-Bereitstellung des Textkorpus ist nicht gestattet, wenn wesentliche Teile von Datenbanken enthalten sind. Zu guter Letzt ist die Online-Bereitstellung auch noch vergütungspflichtig.²⁸

Obwohl im Textkorpus selbst auf eine Quellenangabe verzichtet werden kann, ist eine Quellenangabe erforderlich, wenn das Textkorpus als Ganzes für eine Forschungsgruppe bereitgestellt wird, sofern Werke wiedererkennbar sind.²⁹

Nach der deutschen Umsetzung ist eine Aufbewahrung und Archivierung des Textkorpus nur für Forschungsorganisationen und ihre Forschenden, Bibliotheken und Kulturerbeeinrichtungen und nur so lange erlaubt, wie es „für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich“ ist (Projektbezug). Eine Aufbewahrung darf auch bei externen Einrichtungen erfolgen, wenn der Zugang auf berechnete Forschende bzw. Organisationen beschränkt bleibt. Eine Nachnutzung ist wegen der Beschränkung auf die Zwecke der wissenschaftlichen Forschung und ihrer Überprüfung jedoch nicht gestattet.³⁰

Eine Vergütung ist für die Anfertigung von Kopien und Durchführung des TDM selbst nicht, wohl aber für die Veröffentlichung und Online-Bereitstellung geschuldet. Der Vergütungsanspruch ist verwertungsgesellschaftspflichtig.³¹

23 Brehm, Guidelines, S. 23 ff., 2022.

24 Brehm, Guidelines, S. 15 ff., 2022.

25 Brehm, Guidelines, S. 99 ff., 2022.

26 Brehm, Guidelines, S. 76 ff., 2022.

27 Brehm, Guidelines, S. 79 f., 2022.

28 Brehm, Guidelines, S. 82 f., 2022.

29 Brehm, Guidelines, S. 82, 2022.

30 Brehm, Guidelines, S. 83 f., 2022.

31 Brehm, Guidelines, S. 84 f., 2022.

Bei der Bereitstellung des Ausgangsmaterials sind technische Schutzmaßnahmen der Rechtsinhaberinnen und -inhaber zu beachten. Rechtsinhaberinnen und -inhaber dürfen nach § 60d Abs. 6 UrhG neue Fassung:

- Schutzmaßnahmen zum Schutz der „Sicherheit und Integrität der Netze und Datenbanken durch Vervielfältigungen“ ergreifen, auch wenn dies die Durchführung von TDM für wissenschaftliche Zwecke behindert.
- Nutzungsberechtigungen prüfen.
- Schutzmaßnahmen anbringen, die nicht beseitigt werden dürfen.

Die Schutzmaßnahmen dürfen nicht umgangen werden, müssen andererseits aber verhältnismäßig sein und dürfen die Durchführung des TDM nicht unterbinden. Rechtsinhaberinnen und -inhaber müssen berechtigten Kulturerbeinrichtungen und Forschungsorganisationen die Mittel zur Verfügung stellen, um die Schutzmaßnahmen zu beseitigen und TDM durchführen zu können.³²

Der Gesetzgeber hat Rechtsinhaberinnen und -inhaber bewusst Spielraum eingeräumt, das TDM-Angebot unterschiedlich auszugestalten. Sofern TDM dadurch nicht unterbunden wird, müssen die Angebote der Rechtsinhaberinnen und -inhaber genutzt werden.³³

Allgemeines TDM (§ 44b UrhG)

Seit dem 7.6.2021 gibt es mit der neuen Schrankenregelung § 44b UrhG eine Grundlage, mit der TDM von jedermann zu jedwedem (legalen) Zweck durchgeführt werden darf, also auch zu kommerziellen Forschungszwecken.³⁴ Forschende, die nichtkommerzielle Forschungszwecke verfolgen, können sich aussuchen, auf welche Schrankenregelung sie sich stützen.

Diese Regelung ist abdingbar, so dass Rechtsinhaberinnen und -inhaber die Nutzung ihrer Inhalte für allgemeines TDM verbieten können.³⁵ Sofern die Inhalte digital zur Verfügung gestellt werden, muss das Verbot maschinenlesbar sein.³⁶ Wie auch bei nichtkommerziellem TDM müssen Schutzmaßnahmen der Rechtsinhaberinnen und -inhaber beachtet werden, Selbsthilfe ist nicht erlaubt.³⁷

Der Umfang der Nutzungserlaubnis entspricht dem in § 60d UrhG neue Fassung.³⁸

Bei allgemeinem TDM ist die Online-Bereitstellung, öffentliche Wiedergabe und Archivierung nicht gestattet.

Die Erlaubnis kann jedoch bei Rechtsinhaberinnen und -inhabern erfragt werden.³⁹

Eine Quellenangabe ist nicht erforderlich für die Vervielfältigung, die Aufbereitung der Werke und bei Durchführung des TDM. Da eine Online-Bereitstellung und Veröffentlichung des erstellten Textkorpus nicht gestattet ist, entfällt die Pflicht zur Quellenangabe.⁴⁰

Es besteht eine Löschpflicht: Das Textkorpus muss nach Abschluss des TDM gelöscht werden, wenn es nicht mehr „erforderlich“ ist (§ 44b Abs. 2 S. 2 UrhG). Eine Nachnutzung des erstellten Textkorpus für andere Projekte ist damit ausgeschlossen. Es bleibt jedoch möglich, durch Verweis auf die Quellen und Hinweise zur Aufbereitung des Textkorpus und ggfls. Veröffentlichung von selbst programmierter Software oder Hinweise auf genutzte Tools anderen Projekten die eigenständige Zusammenstellung eines im wesentlichen gleichen Textkorpus zu ermöglichen.

Rolle der Bibliotheken

Bei der Nutzung von wissenschaftlicher Fachliteratur für TDM haben die Bibliotheken eine zentrale Rolle:

Sie stellen den Forschenden die wissenschaftliche Literatur bereit, wollen aber auch Dienstleistungen und Infrastruktur für Forschende bereitstellen. Darüber hinaus wollen sie TDM für die Verbesserung der eigenen Arbeitsprozesse und die Neu- bzw. Weiterentwicklung ihrer Dienstleistungen nutzen. Bibliotheken könnten daher eine zentrale Rolle bei Vermittlung, Aufbereitung und Bereitstellung der Materialien spielen. Darüber hinaus sind sie stark in der Kooperation mit anderen Bibliotheken und Forschenden.

Gleichzeitig sind Bibliotheken Vertragspartner der Rechtsinhaberinnen und -inhaber. Damit sind die Bibliotheken eine wesentliche Schnittstelle: Nur Bibliotheken und Rechtsinhaberinnen und -inhaber kennen den Vertragsinhalt, auf dem aber der Zugriff der Forschenden und die Nutzungserlaubnis für die Bibliothek selbst basiert. Wenn die Verträge vor dem 1.3.2018 geschlossen wurden, ist der Vertragsinhalt entscheidend für die Frage, ob TDM zulässig ist. Erfolgte der Vertragsschluss nach dem 1.3.2018, setzen sich die TDM-Schranken durch, da TDM jedenfalls zu nichtkommerziellen wissenschaftlichen Zwecken nicht mehr vertraglich eingeschränkt werden kann. Aus den

³² Brehm, Guidelines, S. 80 ff., 2022.

³³ Brehm, Guidelines, S. 81 f., 2022.

³⁴ Brehm, Guidelines, S. 85, 2022.

³⁵ Brehm, Guidelines, S. 85, 2022.

³⁶ Brehm, Guidelines, S. 85 f., 2022.

³⁷ Brehm, Guidelines, S. 86, 2022.

³⁸ Brehm, Guidelines, S. 86, 2022.

³⁹ Brehm, Guidelines, S. 86 ff., 2022.

⁴⁰ Brehm, Guidelines, S. 88, 2022.

Verträgen ergibt sich auch, ob ein Risiko besteht, dass Rechtsinhaberinnen und -inhaber Schadensersatzansprüche wegen einer Vertragsverletzung haben und wie groß das Risiko ist, dass Ansprüche vor einem Gericht im EU-Ausland geltend gemacht werden könnten.

Zentrale Frage für die Gestaltung von Dienstleistungen der Bibliotheken bei TDM ist natürlich, ob ein durch Bibliotheken z.B. zu einem bestimmten Fachausschnitt einmal zusammengestelltes und aufbereitetes Textkorpus für weitere Projekte aufbewahrt und zur Verfügung gestellt werden kann?

Oberste Maxime bei der Ausgestaltung von Dienstleistungen ist aus rechtlicher Perspektive, dass im Rahmen einer auf TDM-Schranken basierenden Nutzung keine Parallelarchive entstehen dürfen. Eine anlasslose Herstellung eines Textkorpus ist daher nicht erlaubt. Es muss immer ein Bezug zu einem oder mehreren konkreten Forschungsprojekten bestehen. Eine öffentliche Zugänglichmachung ist nur an einen „bestimmt abgegrenzten Kreis von Personen für gemeinsame wissenschaftliche Forschung“ oder einzelnen Dritten zur „Qualitätsprüfung“ gestattet. Eine Aufbewahrung des Textkorpus ist nur so lange gestattet, wie für den konkreten Zweck des Forschungsprojekts oder die Prüfung der wissenschaftlichen Ergebnisse erlaubt. Im Anschluss muss das Textkorpus gelöscht werden – auch bei der Bibliothek.

Bei der Bereitstellung des Ausgangsmaterials oder des Textkorpus an Forschende müssen die Voraussetzungen der TDM-Schranken bei den Forschenden selbst vorliegen. Sofern Bibliotheken den Zugang zu den Inhalten für TDM vermitteln, liegt die Verantwortung für die „Rechtmäßigkeit des Zugangs“ der Forschenden bei den Bibliotheken. Fazit ist, dass die Bibliotheken projektbezogen die Texte für das Textkorpus, das für TDM genutzt werden soll, aufbereiten dürfen. Sie dürfen die aufbereiteten Textkorpora aber nicht „auf Vorrat“ speichern und auch anderen Projekten zur Verfügung stellen.

Kontrovers wird die Frage diskutiert, wie mit Sprachmodellen umzugehen ist, die im Rahmen des Projekts mit urheberrechtlich geschützten Inhalten trainiert wurden. Dürfen diese Sprachmodelle nur für die konkret mit dem Projekt verfolgten Ziele unter Einhaltung der Voraussetzungen der TDM-Schranken eingesetzt werden? Diese Frage stellt sich insbesondere, wenn die Texte und Daten aus dem Textkorpus nicht in wiedererkennbarer Form im Sprachmodell gespeichert und auch nicht im Rahmen der Benutzung als Ergebnis ausgegeben werden. **I**

Literaturverzeichnis

- Brand, Jonathan: Copyright Implications of the Relationship Between Generative Artificial Intelligence and Text and Data Mining, BLOG infojustice, 27.10.2023, <<https://infojustice.org/archives/45509#:~:text=%E2%80%9CTDM%20is%20a%20crucial%20first,of%20data%20or%20experience%2C%20without>>, [08. Februar 2024].
- Brehm, Elke: Guidelines zum Text und Data Mining für Forschungszwecke in Deutschland, 28.10.2022, Online: <<https://oa.tib.eu/renate/handle/123456789/10352>>.
- Dreier, Thomas / Schulze, Gernot / Specht, Louisa (Hg.), Urheberrechtsgesetz Verwertungsgesellschaftengesetz Kunsturhebergesetz Kommentar, Verlag C. H., 7. Auflage 2022.
- Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes vom 31.05.2021, Bundesgesetzblatt 2021 Teil I Nr. 27 S. 1204, Online: <http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl121s1204.pdf>, [12. Februar 2024].
- Jani, Ole / Vonhien, Maximilian: Gutachten zu bestimmten kollisionsrechtlichen Fragen zur umfassenden Nutzung von lizenzierten Bibliotheksbeständen zum Zwecke des Text und Data Mining, 2022, Online: <<https://oa.tib.eu/renate/handle/123456789/10340>>.
- Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, Amtsblatt der Europäischen Union L130/92, Online: <<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32019L0790&rom=DE>>, [12. Februar 2024].
- Urheberrechtsgesetz vom 9. September 1965 (BGBl. I S. 1273), zuletzt geändert durch Artikel 25 des Gesetzes vom 23. Juni 2021 (BGBl. I S. 1858), <<https://www.gesetze-im-internet.de/urhgf/>>, [12. Februar 2024].

Dieser Artikel wird unter der Lizenz CC by 3.0 Deutschland (Kurzfassung und vollständiger Lizenztext: <https://creativecommons.org/licenses/by/3.0/de/deed.de>) veröffentlicht. Die Lizenz gilt nicht für Zitate oder referenzierte Quellen.

.....

Elke Brehm

Fachreferentin für Rechtswissenschaften, stv. Justiziarin, Datenschutzbeauftragte der TIB, Bereich Wissenschaftlicher Dienst, Programmbereich A - Bestandsentwicklung und Metadaten

<https://orcid.org/0000-0001-8224-7047>

Elke.Brehm@tib.eu

.....