
» Themen

• » [Über uns](#)

Data Mining 4. Februar 2015 | [Christoph Bruch](#)

Wer schürfen kann, ist klar im Vorteil



Wer ein Buch besitzt, muss nicht erneut um Erlaubnis fragen, um es zu lesen. Wenn Computer in Datenbergen nach Informationen schürfen, sollte der gleiche Grundsatz gelten. Lizenzen für Data Mining behindern Forschung und Innovation in Europa, kommentiert Christoph Bruch von der Helmholtz-Gemeinschaft.

Sollen Rechteinhaber urheberrechtlich geschützter Werke das Text- und Data-Mining kontrollieren dürfen? Dabei geht es um die systematische, computergestützte Analyse der Inhalte dieser Werke. Wohlgemerkt: Um Werke, auf die legal Zugriff besteht.

Diese Frage ist ein zentraler Streitpunkt der laufenden Verhandlungen um eine Novellierung der EU-Urheberrechts-Richtlinie. Die deutsche Abgeordnete Julia Reda hat die Frage in [ihrem Entwurf für das Europäische Parlament](#) aufgegriffen.

Wir alle nutzen über Suchmaschinen die Ergebnisse von Text- und Data-Mining tagtäglich. In einzelnen Wissenschaftsfeldern spielen spezialisierte Formen von

Text- und Data-Mining bereits eine wichtige Rolle. Für Unternehmen, die entsprechende Dienste anbieten, eröffnet sich ein neuer Markt. Die Verleger wissenschaftlicher Publikationen möchten diese Auswertungen jedoch kontrollieren und hemmen damit Wissenschaft und Wirtschaft.

Alte EU-Kommission tappte in die Lizenzfalle

In der letzten EU-Kommission fanden sie offene Ohren für ihre Interessen. Mit den Rahmenbedingungen der unter dem Titel „[Licenses for Europe](#)“ organisierten Verhandlungen hatte die Kommission die von den Verlegern propagierte Regelung bereits als Ergebnis nahegelegt. Die Verwerter-Industrie proklamierte, Text- und Data-Mining mit Hilfe von Lizenzen allen Interessierten zu fairen Konditionen ermöglichen zu können.

Demgegenüber vertraten Nutzerverbände einschließlich der Netzaktivisten, der Bibliotheken und der Wissenschaft die Ansicht, dass Text- und Data-Mining eine dem Lesen nachgeordnete Nutzung ist – und deswegen weder das Mining selbst noch die Verbreitung der Analyseergebnisse durch das Urheberrecht beschränkt werden sollen. Sie brachten diese Forderung auf die Formel: „The right to read is the right to mine!“

Als sich abzeichnete, dass allein die Teilnahme der Nutzervertreter an den Gesprächen dazu führen könnte, inakzeptable Ergebnisse zu legitimieren, verließen die Nutzervertreter [unter Protest](#) die Verhandlungen.

Data Mining EU-weit nicht eigens geregelt

Die Behauptung der Verwerter-Industrie, nicht lizenziertes Text- und Data-Mining verletze ihre Urheberrechte, ist erklärungsbedürftig. Denn Text- und Data-Mining wird weder in der europäischen [Urheberrechts-Richtlinie](#) noch in einem der darauf beruhenden nationalen Urheberrechtsgesetze als geschütztes Verwertungsrecht der Urheber erwähnt.

Ein urheberrechtlich begründetes Verbot der Nutzung dieser Techniken muss daher an der Verletzung anderer geschützter Verwertungsrechte festgemacht werden. Nach gängigem Verständnis ist dies das Vervielfältigungsrecht. Dabei wird unterstellt, dass Text- und Data-Mining die Herstellung von Kopien einschließt.

Das Unbehagen gegenüber dem Kontrollanspruch der Verwerter wird noch einmal verstärkt, wenn man sich klar macht, dass Text und Datenmining

- auf das urheberrechtlich nicht geschützte Lesen der zu analysierenden Inhalte aufbaut und
- darauf zielt, den informativen Gehalt eines Werkes zu erfassen, der ebenfalls durch das Urheberrecht nicht geschützt ist.

In anderen Ländern genehmigungsfrei

Im internationalen Vergleich sind Nutzer in Europa hinsichtlich des Text- und

Data-Mining benachteiligt. In den USA ist basierend auf der [Fair-use-Doktrin](#) Text- und Data-Mining im nicht-kommerziellen wissenschaftlichen Kontext genehmigungsfrei möglich. In Japan ermöglicht dies eine explizite Ausnahmeregelung (Schranke) und seit einigen Monaten gilt auch in Großbritannien solch eine Schranke.

Diese Ansätze teilen jedoch den problematischen Aspekt, Text und Data-Mining lediglich für nicht-kommerzielle Zwecke zu ermöglichen und damit die Entwicklung entsprechender kommerzieller Dienste zu hemmen.

Diese Problematik wird noch einmal deutlicher, wenn man berücksichtigt, dass zum Beispiel große Suchmaschinenbetreiber aufgrund ihrer Gatekeeper-Funktion gegenüber Rechteinhabern eine so machtvolle Position haben, dass diese ihnen kostenfrei die benötigte Lizenz einräumen.

Forschung und Innovation werden ausgebremst

Eine Prognose [des McKinsey „Global Institute“](#) illustriert die wirtschaftliche Dimension der Nutzung von „Big Data“ und Text- und Data-Mining. Bei der Gesundheitsversorgung sieht sie jährliche Effizienzgewinne in Europa in Höhe von 100 Milliarden US-Dollar, in den USA in Höhe von 300 Milliarden.

Die Bedeutung von Text und Data-Mining für Effizienzsteigerungen oder Innovation hängt aber nicht zwingend an Big-Data-Szenarien, wie sie von McKinsey untersucht wurden. In der Wissenschaft wird Text- und Data-Mining oft in viel kleineren Dimension und trotzdem sehr innovativ eingesetzt.

Eine Arbeitsgruppe am Berliner Max-Delbrück-Zentrum für molekulare Medizin sucht mit Hilfe von Text- und Data-Mining in wissenschaftlichen Publikationen nach dem Zusammenhang von Krankheiten mit bestimmten Gensequenzen. Aufgrund der urheberrechtlichen Situation sind diese Recherchen auf die Abstracts der Publikationen beschränkt. Der medizinische Fortschritt wird gebremst.

Reda-Entwurf für Ausnahmeregelung

Der jüngst im Europäischen Parlament vorgestellte [Berichtsentwurf von Julia Reda \(PDF\)](#) mit Vorschlägen zur Novellierung der Urheberrechts-Richtlinie sieht vor, eine für alle EU-Mitgliedstaaten verbindliche Urheberrechtsausnahme für Text- und Data-Mining einzuführen, die auch die kommerzielle Nutzung einschließt. Die Realisierungschancen für den Vorschlag einer Schrankenregelung sind derzeit nicht absehbar.

Der deutsche Gesetzgeber ist beim Urheberrecht in Brüssel zuletzt mit der Einführung des [Leistungsschutzrechtes für Presseverleger](#) negativ aufgefallen, weil solche nationalen Vorstöße der von der Europäischen Kommission angestrebten Harmonisierung des Urheberrechtes entgegenstehen.

Jetzt kommt es darauf an, die Bundesregierung davon zu überzeugen, im Kontext der angestrebten europäischen Novellierung des Urheberrechts auf die

Unterstützung von Innovation zu setzen. Diesem Interesse dient die Text- und Data-Mining-Schranke, die jetzt von Julia Reda in die Diskussion gebracht wurde.



Dr. Christoph Bruch ist für das [Open-Science-Koordinationsbüro der Helmholtz-Gemeinschaft](#) tätig. Er studierte Politikwissenschaft in Frankfurt/M. und Berlin, berufliche Stationen an der FU Berlin, dem Deutschen Institut für Urbanistik und der Max-Planck-Gesellschaft. Zu seinen Arbeitsschwerpunkten gehört die Formulierung von nationalen und internationalen Policies, die den Übergang zu Open Access und Open Science unterstützen sollen.

Zum Thema bei iRights.info

- [Information Mining: Aus dem Steinbruch der Wissenschaft](#)

[Big Data in der Medizin](#)

[Berichtsentwurf fordert einheitlicheres EU-Urheberrecht](#)

Zum Thema im Internet

- [Europäisches Parlament: Entwurf eines Berichts über die Umsetzung der Richtlinie 2001/29/EC \(PDF\)](#)

3 Kommentare

- [1 Charlotte Schubert](#) am 29. Juli, 2015 um 14:10

Lieber Herr Bruch, sind Sie sicher, daß nach dem amerikanischen Fair Use das Textmining im wiss. Kontext genehmigungsfrei ist? Wenn ja, würden Sie dazu Literatur oder Rechtsprechungshinweise geben können?

Vielen Dank und Grüße

Ch.Schubert

- [2 Christoph Bruch](#) am 30. Juli, 2015 um 06:41

Sehr geehrte Frau Schubert,

herzlichen Dank für Ihr Interesse am Thema Urheberrecht und TDM. Den erbetenen Literaturhinweis finden Sie unten. Bitte berücksichtigen Sie, dass die Anwendung der Fair-Use –Doktrin grundsätzlich mit Unsicherheiten

verbunden ist, da ihre Anwendbarkeit auf bestimmte Nutzungsszenarien von Einzelfallentscheidungen abgeleitet wird.

Gruß

Christoph Bruch

Cox, Krista L. (5. VI. 2015): Text and Data Mining and Fair Use in the United States, Association of Research Libraries.

<http://www.arl.org/storage/documents/TDM-5JUNE2015.pdf>

ARL has released an “Issue Brief: Text and Data Mining and Fair Use in the United States” (PDF), which describes the role and usefulness of text and data mining, provides a short background of fair use, and presents an analysis of fair use in text and data mining, including eight cases that support fair use in this context.

No researcher can read all relevant research articles that are published in her field of interest. Even if she could, she would not be able to detect patterns in the research results that emerge only from large-scale computational analysis, known as text and data mining (TDM). Researchers who want to perform TDM on copyrighted research articles might seek clarity about whether they need permission from journal publishers or whether copyright’s fair use doctrine permits TDM on accessible articles. **In almost all cases, performing TDM on accessible articles is a fair use.**

TDM almost always involves copying, but not all copying amounts to copyright infringement. Numerous courts in the United States have upheld the reproduction necessary to perform TDM as fair use, even though the content being copied into the database is copyrighted. Fair use is a flexible limitation and exception that allows copyright law to adapt to changing circumstances and new technologies and helps ensure a balanced copyright system. Thus, while the United States does not have a specific limitation or exception to explicitly allow TDM, fair use has accommodated the creation and growth of TDM as a new research tool.

TDM may be used for a variety of purposes, some of which are explicitly referenced in Section 107 of the US Copyright Act, such as scholarship and research. Beginning with a 2003 case involving the incorporation of images in a search engine, in at least eight different cases, courts have found that the creation of a database for TDM and its use amounts to fair use. The purposes have ranged from research by scholars, to use by politicians, to checking for plagiarism. Many of these courts have focused heavily on the benefit that TDM provides to the public, because they “enhanc[e] information-gather techniques.” In assessing the four factors of fair use purpose and character of the use, nature of the copyrighted work, amount and substantiality of portion used, and effect on the potential market courts have emphasized the transformative nature of searchable databases and TDM, noting the

unlikelihood of adverse impact on the original market for the work, and upheld fair use.

While this issue brief covers fair use and TDM in the United States, TDM is an issue of concern in other countries, as well. Internationally, 171 organizations including ARL, have called for the removal of barriers with respect to data, through the Hague Declaration on Knowledge Discovery in the Digital Age. The Hague Declaration calls for clarity around the scope of intellectual property law as well as calling for better infrastructure to allow for content mining.

- [3 Charlotte Schubert](#) am 10. August, 2015 um 09:37

Lieber Herr Bruch, ich danke Ihnen sehr herzlich, dieses Statement der ARL ist außerordentlich nützlich und sollte möglichst großflächig bekannt gemacht werden!

mbG

Ch.Schubert

Was sagen Sie dazu?

 Name **Mail** **Website**

Alle originären Inhalte auf dieser Website sind, soweit nichts anderes vermerkt ist, urheberrechtlich geschützt und lizenziert unter der [Creative Commons Namensnennung 4.0 International Lizenz](#).

» [Impressum / Datenschutz](#)