

Helmholtz Open Access Koordinationsbüro

Rechtliche Aspekte von Text und Data Mining

Begriffserläuterung

Mit dem Begriff Text und Data Mining (im Weiteren Content Mining) wird eine große Bandbreite von Techniken bezeichnet, die der computergestützten Suche in sowie der Analyse und Vernetzung von Daten und Texten dienen. Die Anwendung dieser Techniken ist nicht auf bestimmte Inhalte und Formate beschränkt. Aus diesem Grund wird teilweise auch allgemein von Content Mining gesprochen. Content Mining muss nicht auf umfangreiche Korpora angewendet werden, gerade hier entfalten die Techniken jedoch ihre besondere Stärke.

Motivation für die Nutzung von Content Mining Techniken

Die für einzelne WissenschaftlerInnen nicht mehr überschaubare Menge an Fachveröffentlichungen und Forschungsdaten macht den Einsatz von Content Mining schon heute oft unverzichtbar. Viele WissenschaftlerInnen nutzen diese Techniken indirekt, z. B. über Suchmaschinen oder bibliographische Datenbanken ohne dies bewusst als Content Mining aufzufassen. Content Mining kann weit über die Nutzung von Suchmaschinen hinausgehen und z. B. eigens entwickelte Softwaretools für spezielle semantische Analysen nutzen. Eine stetig steigende Anzahl von WissenschaftlerInnen betreiben eigenständig Content Mining und benötigen dafür umfangreichen Zugriff auf Forschungsdaten und Publikationen.

Content Mining wird als Motor zur Steigerung der Leistungsfähigkeit der Wissenschaft gesehen. Dies gilt sowohl für die Erlangung von Erkenntnissen, die ohne ihren Einsatz nicht erzielbar sind, als auch hinsichtlich der Beschleunigung von Forschungsprozessen.

Die Möglichkeit der Nutzung dieser Techniken und der Zugriff auf die jeweils interessierenden Daten und Publikationen beeinflussen unmittelbar die Wettbewerbsfähigkeit von Wissenschaftlern und damit auch die ihrer Institution sowie – bei unterschiedlicher Rechtslage – ihres Landes.

Rechtsanspruch der Rechteinhaber

Vor diesem Hintergrund ist es problematisch, dass wissenschaftliche Verlage für sich in Anspruch nehmen, die Nutzung der von Ihnen publizierten Inhalte für Content Mining sei genehmigungspflichtig. Die oben skizzierten Analysen seien demnach nur in einem von den Verlagen genehmigten Umfang möglich und für die Einräumung dieses Rechtes müsste mutmaßlich bezahlt werden.

Die Rechteinhaber begründen ihre Ansprüche mit Verweisen auf

- das Urheberrecht,
- das Datenbankschutzrecht sowie mit
- Verboten oder Beschränkungen die in Subskriptionsverträgen enthalten sind.

Dieser Anspruch ist hinsichtlich des Urheberrechtes erklärungsbedürftig, weil Daten und Fakten, das Objekt von Content Mining, im Regelfall keinen urheberrechtlichen Schutz genießen. Ebenso ist das Lesen, die Nutzungsform auf der das Content Mining aufbaut, nicht urheberrechtlich geschützt.

Das Urheberrecht gibt den Urhebern Kontrolle über eine Reihe konkret im Gesetz aufgelisteter Nutzungsformen. Content Mining ist dort nicht genannt. Die Rechteinhaber verweisen deshalb auf das

Zitationsvorschlag: Rechtliche Aspekte von Text und Data Mining. Helmholtz Open Science Briefing. Stand 14.10.2013. Version 1.0. <http://oa.helmholtz.de/index.php?id=68#c2368>

Impressum: Herausgegeben vom Helmholtz Open Access Koordinationsbüro. Redaktion: Dr. Christoph Bruch, Heinz Pampel und Dr. Paul Schultze-Motel. Kontakt: open-access@helmholtz.de. Lizenz: Creative Commons Namensnennung 3.0 Deutschland (CC BY 3.0 DE). Stand: 14.10.2013.

Vervielfältigungsrecht. Sie reklamieren, dies würde durch elektronische Kopien, die im Zuge des Content Mining erstellt werden, tangiert. Vom Vervielfältigungsrecht ausgenommen sind jedoch sogenannte temporäre Kopien, die im Kontext der Datenverarbeitung oder Weiterleitung entstehen. Ob und welche Form des Content Mining genehmigungspflichtig ist, wird damit zu einer komplizierten Frage, deren Beantwortung von Details der angewendeten Technik abhängt.

Content Mining benötigt die Zugriffsmöglichkeit auf die zu durchsuchenden/analysierenden Korpora. Hier greift der Datenbankschutz, der das genehmigungsfreie Herunterladen von Inhalten stark einschränkt. Dieses Leistungsschutzrecht ist eine europäische Besonderheit.

Soll Content Mining auf subskribierte Inhalte angewendet werden, greifen zusätzlich zur gesetzlich begründeten Rechtslage mögliche Regelungen zu Nutzungsbeschränkungen des Content Mining in Subskriptionsverträgen.

Politische Reaktion

Die Unzufriedenheit der Wissenschaft mit der aktuellen Rechtslage hinsichtlich des Content Mining ist ein Kritikpunkt unter vielen, die den Ruf nach einer Novellierung der europäischen Copyright Richtlinie (29/2001) begründen. Die Rechteinhaber halten dem entgegen, Standardlizenzen für Content Mining würden eine Novellierung überflüssig machen. Für Wissenschaftsorganisationen ist es wichtig, ihre grundsätzliche Interessenvertretung für ein wissenschaftsfreundlicheres Urheberrecht nicht aufzugeben, auch wenn im Zeitraum vor der angestrebten Änderung vertragliche Lösungen akzeptiert werden.

Zu den politischen Reaktionen zählt auch die Förderung des Open-Access-Publizierens, denn es beschränkt die Kontrollmacht der Verlage, wenn die Publikationen mit einer Lizenz versehen werden, die alle legitimen Nutzungsformen unter der Voraussetzung der Nennung der Urheber frei geben. Im Idealfall sollten Wissenschaftsorganisationen darauf bestehen, Publikationsgebühren ausschließlich für Publikationen mit CC-BY-Lizenz zu zahlen.

Open Access-Publizieren reicht nicht aus, um eine genehmigungs- und kostenfreie Nutzung der Werke z. B. für Content Mining sicher zu stellen. Zusätzlich wird eine Infrastruktur benötigt, über die die Zugriffsmöglichkeit auf die Werke gewährleistet wird. Eine Lizenz wie CC-BY regelt die Nutzungsmöglichkeiten, wenn der Zugriff gegeben ist, sie schließt keine Pflicht zur Zugänglichmachung ein. Die Zugänglichmachung der Werke wird über Datenbanken realisiert. Für diese Sammlungen gilt in Europa unabhängig vom Rechtsschutz für die einzelnen Werke ein eigenes Datenbankschutzrecht.

Konsequenzen für WissenschaftlerInnen

Von WissenschaftlerInnen kann unter den rechtlichen Umständen nicht erwartet werden, die Rechtmäßigkeit der jeweils zum Einsatz kommenden Content-Mining-Technik zu beurteilen. Ein Verzicht auf Content Mining ist aus der Perspektive der betroffenen WissenschaftlerInnen jedoch keine akzeptable Option, nicht nur weil sie nicht auf die mit Content Mining verbundene Analysefähigkeit verzichten möchten, sondern auch weil z. B. ihre Konkurrenten in den USA aufgrund einer differierenden Rechtslage nicht in dem Maße wie ihre europäischen Kollegen an der Nutzung der Technik gehindert werden. Deshalb liegt die Vermutung nahe, dass vielfach WissenschaftlerInnen, möglicherweise auch an Helmholtz-Zentren, Content Mining ohne eine Berücksichtigung der rechtlichen Umstände betreiben.

Handlungsoptionen

Aus der Problematik ergeben sich für Wissenschaftsorganisationen eine Reihe von Handlungsoptionen:

- Verzicht auf Content Mining
- stärkere Kontrolle der Mitarbeiter zur Vermeidung von Störerhaftung.
- Unterstützung der Mitarbeiter durch Aufklärung und Rechtsberatung.
- Verweigerung des Abschlusses von Subskriptionsverträgen mit restriktiven Regelungen zum Content Mining, bzw. Versuch der Durchsetzung einer Regelung im Sinne der Wissenschaft. Für Letzteres kann eine Musterklausel dienen, die im Rahmen der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen erarbeitet wurde.
- Information über die Bedeutung eines zukünftigen wissenschaftsfreundlichen Urheberrechtes.

Empfehlungen

Kein Verzicht auf Content Mining

Ein Verzicht auf Content Mining ist aus der Perspektive von Wissenschaftsorganisationen aus denselben Gründen keine Option aus welchen er auch aus der Perspektive der einzelnen WissenschaftlerInnen keiner ist.

Keine stärkere Kontrolle der MitarbeiterInnen

Bisher sind keine Fälle bekannt, in welchen illegales Content Mining durch WissenschaftlerInnen zu rechtlichen Konsequenzen geführt hat. Es kommt allenfalls vereinzelt zu Schwierigkeiten, weil durch umfangreiches Herunterladen von Artikeln mit Verlagen vereinbarte Lizenzen (etwa Quoten) verletzt wurden und infolgedessen der Zugang zu einer Datenbank vorübergehend gesperrt wurde. Diese Quotenüberschreitungen stehen jedoch nicht notwendig in einem Zusammenhang mit Content Mining. Diese Fälle lassen sich in der Regel informell klären.

Aufklärung der MitarbeiterInnen über die Problematik

Solch eine Information dient zwei Zwecken:

- a) einer Verbesserung der Möglichkeit der WissenschaftlerInnen, eine informierte Entscheidung zu treffen
- b) dem Aufbau von Unterstützung innerhalb der Organisationen für ein entsprechendes Verhandlungsmandat beim Abschluss von Subskriptionsverträgen.

Verweigerung der Akzeptanz von restriktiven Regelungen zum Content Mining in Subskriptionsverträgen

Es sollte sichergestellt werden, dass bei Vertragsverhandlungen mit Verlagen etwaige Regelungen zum Content Mining genau geprüft bzw. eigene Regelungsvorschläge präferiert werden. Die Wahrscheinlichkeit zur Durchsetzung restriktiver Regelungen hängt nicht nur vom Verhandlungsmandat der Vertreter der Wissenschaftsorganisationen, sondern wesentlich von deren Rückhalt in der Organisation ab.

Information über die Bedeutung eines zukünftigen wissenschaftsfreundlichen Urheberrechtes auf nationaler und Europäischer Ebene

Die aktuelle und eine möglicherweise aus der Perspektive der Wissenschaftsorganisationen verschlechterte Rechtslage perpetuiert die Übermacht der Verlage und die damit einhergehenden Beschränkungen und Kosten. Die Wissenschaftsorganisationen sollten deshalb bestehende Aktivitäten zur Verdeutlichung der Bedürfnisse der Forschung beibehalten oder verstärken – auf nationaler Ebene über die Allianz der deutschen Wissenschaftsorganisationen, auf europäischer Ebene über Science Europe.