Text Mining öffentlich zugänglicher Daten

Smart Data Begleitforschung,
Fachgruppenworkshop

Benjamin Bremert,
Harald Zwingelberg

Berlin, 4. Mai 2017

Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein

Was ist Text Mining?

- Sowohl Data als auch Text Mining beschreibt die **Analyse von Daten und Suche nach (wirtschaftlich nutzbaren) Mustern.**

- Im Unterschied zu Text Mining, bei dem es um die Analyse von unstrukturierten Daten geht, liegen die Daten beim Data Mining häufig in Datenbanken und insoweit strukturiert vor. Struktur bei Text Mining lediglich implizit.

Rechtliche Herausforderungen

- Schuldrecht / Sachenrecht
  - Beschränkung der Nutzbarkeit und des Crawlens von Datenquellen
- Datenschutzrecht
  - Verarbeitung personenbezogener Daten während des Crawlens und der Analyse
- Urheberrecht
  - Vervielfältigung von Inhalten durch das Crawling und die Analyse

Schuldrecht / Sachenrecht

- Vertragliche Ansprüche gegen Crawling?
  - Wohl nur bei Anmeldung auf einer Plattform oder Nutzung einer API und damit zusammenhängender Anmeldung.

- Unterlassungsanspruch gegen Betreiber des Crawlers?
  - §§ 903, 1004 BGB (aus Servereigentum)
  - § 858 BGB (aus Besitz des angemieteten Servers)
    (P) Zufällige Ergebnisse je nach technischer Infrastruktur
  - §§ 903, 1004 analog BGB (aus analoger Anwendung der Eigentumsvorschriften auf Betreiber von Internetseiten)

Datenschutzrecht

Rechtsgrundlage zur Verarbeitung öffentlich zugänglicher Daten?

- Einwilligung (-)
- Gesetzliche Grundlage:
  - **§ 28 Abs. 1 Satz 1 Nr. 3 BDSG**
  - **§ 28 Abs. 2 Nr. 3 BDSG**
  - **Art. 6 Abs. 1 Satz 1 Lit. f DSGVO**

Datenschutzrecht

§ 28 Abs. 1 Satz 1 Nr. 3 BDSG

„Das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung als Mittel für die Erfüllung eigener Geschäftszwecke ist zulässig … wenn die Daten allgemein zugänglich sind oder die verantwortliche Stelle sie veröffentlichen dürfte, es sei denn, dass das schutzwürdige Interesse des Betroffenen an dem Ausschluss der Verarbeitung oder Nutzung gegenüber dem berechtigten Interesse der verantwortlichen Stelle offensichtlich überwiegt."

- Allgemein zugängliche Daten

  Jedenfalls solche Websites, die allgemein zugänglich sind. Umstritten, ob anmeldepflichtige Websites noch allgemein zugängliche Daten enthalten.

- Berechtigung der Veröffentlichung
- Interessenabwägung

  Verarbeitung ist nur dann unzulässig, wenn das Interesse des Betroffenen an einem Ausschluss der Verarbeitung offensichtlich überwiegt.

Datenschutzrecht

§ 28 Abs. 2 Nr. 3 BDSG

„Die Übermittlung oder Nutzung für einen anderen Zweck ist zulässig ... wenn es im Interesse einer Forschungseinrichtung zur Durchführung wissenschaftlicher Forschung erforderlich ist, das wissenschaftliche Interesse an der Durchführung des Forschungsvorhabens das Interesse des Betroffenen an dem Ausschluss der Zweckänderung erheblich überwiegt und der Zweck der Forschung auf andere Weise nicht oder nur mit unverhältnismäßigem Aufwand erreicht werden kann."

- Wissenschaftliche Forschung

  Der nach Form und Inhalt ernsthafte Versuch zur Ermittlung der Wahrheit durch die planmäßige und zielgerichtete Suche nach neuen Erkenntnissen.
- Wissenschaftliches Interesse überwiegt Interesse an Ausschluss der Verarbeitung erheblich
- Zweck der Forschung sonst nur mit unverhältnismäßigem Aufwand erreichbar.

Datenschutzrecht

Art. 6 Abs. 1 Satz 1 Lit. f DSGVO

„Die Verarbeitung ist nur rechtmäßig, wenn mindestens eine der nachstehenden Bedingungen erfüllt ist: ... die Verarbeitung ist zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt."

- Berechtigtes Interesse des Verantwortlichen oder eines Dritten

  Rechtliche, wirtschaftliche und ideelle Interessen. Normative Wertung, ob Interesse gegen die Rechtsordnung verstößt und kein weniger eingriffsintensives Mittel zur Zweckerreichung zur Verfügung steht.

- Abwägung mit Interessen betroffener Person

  (P) *Kinder als betroffene Person sind überwiegend Schutzwürdig*

Datenschutzrecht

Abwägungsmaßstab

- Unübersehbar keine Verarbeitung gewünscht
- Art der Daten
  - **(P)** Besondere Kategorien personenbezogener Daten, Art. 9 Abs. 1 DSGVO bzw. § 3 Abs. 9 BDSG
  - Aber: Art. 9 Abs. 2 Lit. e DSGVO (Daten von betroffener Person öffentlich gemacht) bzw. § 28 Abs. 6 Nr. 2 BDSG
- Inhalt der Daten
  - (P) genügt sofortiges Verwerfen von Informationen
- Aufgaben und Zweck der Datenverarbeitung
  - **(P)** Erstellung eines Persönlichkeitsprofils

Datenschutzrecht

Informationspflichten

- Grds. §§ 33 Abs. 1 Satz 1 und 2 BDSG
- Ausnahme:
  - § 33 Abs. 2 Satz 1 Nr. 8a BDSG (Zur geschäftsmäßigen Übermittlung, aus allg. zugänglichen Quellen)
  - § 33 Abs. 2 Nr. 5 BDSG (Wissenschaftliche Forschung)
  - § 33 Abs. 2 Nr. 7a BDSG (Für eigene Zwecke, aus allg. zugänglichen Quellen)
- Grds.: Art. 14 DSGVO
- Ausn.: Art. 14 Abs. 5 DSGVO

Dazu gleich

Urheberrecht

Urheberrechtswidrige Vervielfältigungshandlung durch den Crawler?

- Öffentliche Zugänglichmachung, § 19a UrhG

  „Das Recht der öffentlichen Zugänglichmachung ist das Recht, das Werk drahtgebunden oder drahtlos der Öffentlichkeit in einer Weise zugänglich zu machen, dass es Mitgliedern der Öffentlichkeit von Orten und zu Zeiten ihrer Wahl zugänglich ist."

- Bearbeitung, § 23 UrhG

  „Anpassung des Werkes für andere Nutzungsformen oder Übertragung in anderen Werkstoff."

- Sui generis Schutz für Datenbanken, §§ 87a UrhG

  Systematische und Wiederholte Entnahme. Scheitert an restriktiver Auslegung von Art. 7 Abs. 5 Datenbank-RL durch den EuGH, der Rekonstruierung von wesentlichen Teilen erfordert.

Urheberrecht

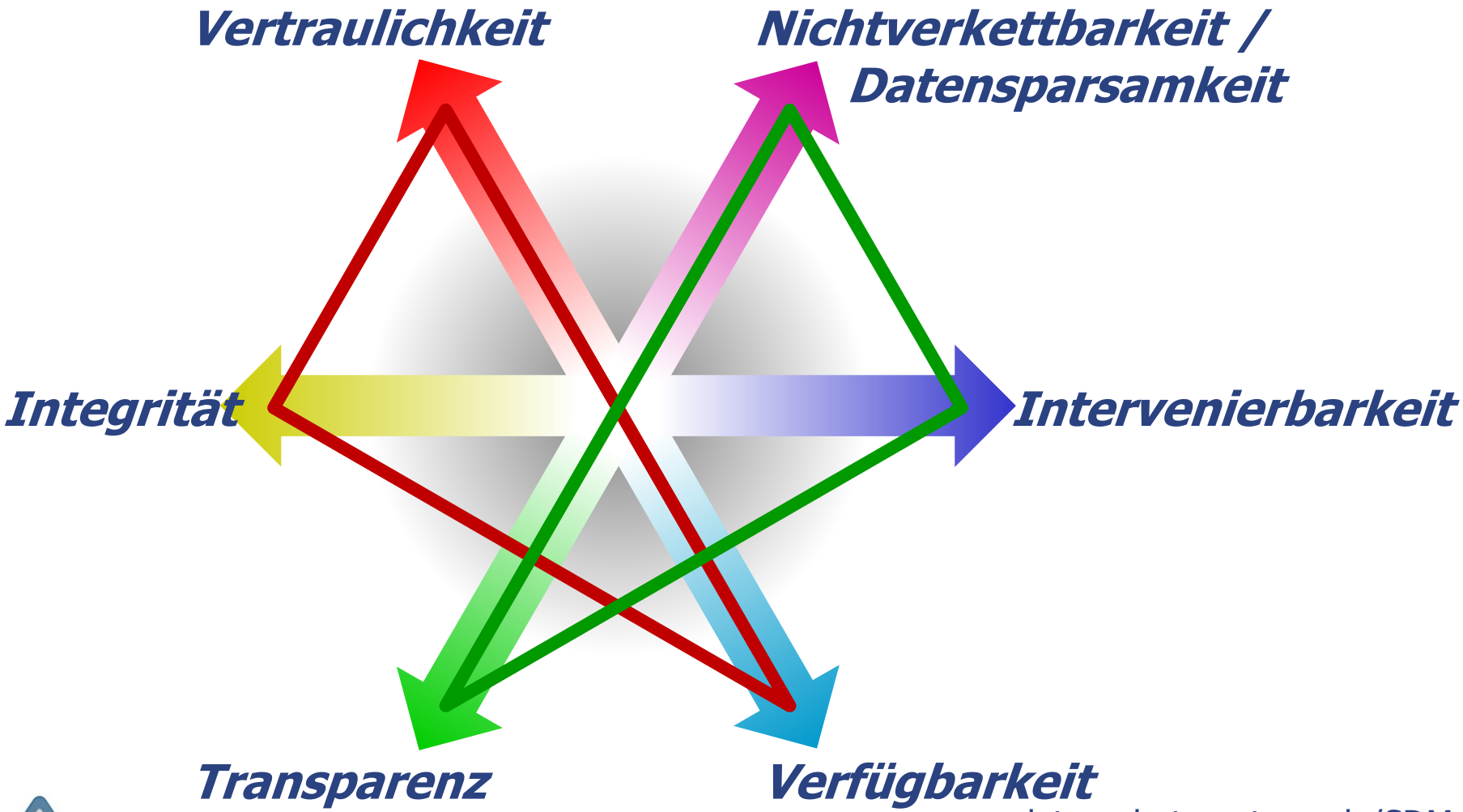
- Vervielfältigung, § 16 UrhG

  „Das Vervielfältigungsrecht ist das Recht, Vervielfältigungsstücke des Werkes herzustellen, gleichviel ob vorübergehend oder dauerhaft, in welchem Verfahren und in welcher Zahl."

Aber: § 44a UrhG

*„Zulässig sind **vorübergehende Vervielfältigungshandlungen**, die **flüchtig** oder **begleitend** sind und einen **integralen und wesentlichen Teil eines technischen Verfahrens** darstellen und deren alleiniger Zweck es ist, … eine **rechtmäßige Nutzung** … eines Werkes oder sonstigen Schutzgegenstands zu ermöglichen, und die **keine eigenständige wirtschaftliche Bedeutung** haben."*

Standarddatenschutzmodell (SDM) vergl. SDBF-workshop Sept. 2016



www.datenschutzzentrum.de/SDM

Gewährleistungsziel Transparenz

Transparenz bezeichnet die Anforderung, dass sowohl Betroffene, als auch die Betreiber von Systemen sowie zuständige Kontrollinstanzen erkennen können, welche Daten für welchen Zweck in einem Verfahren erhoben und verarbeitet werden, welche Systeme und Prozesse dafür genutzt werden, wohin die Daten zu welchem Zweck fließen und wer die rechtliche Verantwortung für die Daten und Systeme hat.

Quelle: Standarddatenschutzmodell, S. 13.

Transparenz bei bestehendem Kundenkontakt

[Exkurs vom Vortrag zu Datamining öffentlicher Daten]

- In der Regel eine Erhebung beim Betroffenen => Art. 13
  - Kundenkontakt, daher Einwilligung möglich
  - Widerruf einer Einwilligung ist zu beachten

- Bei Datenweitergabe an Dritte oder Zweckänderung durch Gewinnung neuer Informationen durch Analyse neue Einwilligung und Information erforderlich.

⇒ Einfache, vereinheitlichte Kommunikationsmöglichkeit mit Betroffenen wünschenswert

Transparenz bei öffentlich zugänglichen Daten

- Erhebung ohne Wissen der Betroffenen und ohne Kontakt, Beispiel iTESA mit Datenerhebung aus dem Netz

- Gesetzliche Rechtsgrundlage: Möglich hier Art. 6 (1) (f) DSGVO zur Wahrnehmung berechtigter Interessen des Verantwortlichen. Folgen:
  ⇒Art. 14 DSGVO Information des Betroffenen
  ⇒Art. 21 DSGVO Widerspruchsrecht – " –

- Öffentliche Angaben z.B. auf der Webseite als Minimum notwendig, Art. 14 (5) (b) am Ende DSGVO.

Ausnahmen zugunsten von Big Data?

- Ausnahme des Art. 14 (5) (b) für Big Data?
  - Auskunftspflicht entfällt, sofern „sich die Erteilung […] als unmöglich erweist oder einen unverhältnismäßigen Aufwand erfordern würde […]"
    Regelbeispiele dieser Ausnahme: insbesondere für öffentliche Archive, Forschungszwecke und Statistik.
  - Vorab: Entgegen reinem Gesetzestext ist eine **Abwägung** zwischen Aufwand für Verarbeiter und Interesse an der Information der betroffenen Person zwingend.
    [so zutreffend Kühling/Buchner/Bäcker Art. 14 DSGVO Rn. 55; schon zu Art. 11 DS-RiLi Dammann/Simitis Art. 11 RiLi Rn. 5]
  - Abwägungserfordernis folgt auch aus auch EU-weit allgemein gültigen Verhältnismäßigkeitsprinzip.

Ausnahmen zugunsten von Big Data?

- Ausnahme des Art. 14 (5) (b) für Big Data?
  - Wird Big Data per se bei vielen und nicht näher bekannten Betroffenen privilegiert?
  - Generelle Ausnahme für Big Data? Wird teilweise vertreten, u.a. mit dem Argument, man könne der Norm keine Untergrenze im Sinne einer Anzahl Betroffener entnehmen, daher sogar bei beliebig kleiner Zahl Betroffener.

    ⇒Aber Abwägung zwingend (s.o.), keine generelle Lösung für Big Data
  - Gedanke des Art. 11 DSGVO keine Identifizierung allein zu Zwecken der DSGVO-Compliance, aber nach Erwägungsgrund 57 sind von Betroffenen beigebrachte Informationen zu Berücksichtigen

    ⇒ ergo: Prozess muss ohnehin etabliert sein

    ⇒ Berücksichtigung bei der Abwägung u.a. bei Quantifizierung des Risikos

Ausnahmen zugunsten von Big Data?

- Ausnahme des Art. 14 (5) (b) für Big Data?
  - Einfluss des verfolgten Zwecks?
  - Ja, siehe Regelbeispiele dieser Ausnahme: insbesondere für öffentliche Archive, Forschungszwecke und Statistik. Hier ist Abwägung entbehrlich bzw. fällt im Regelfall zugunsten des Verarbeiters aus. Vergleichbare Zwecke können ebenfalls privilegiert sein.
  - Daten aus öffentliche Quellen: spricht für Privileg
  - Risikoreiche Verarbeitung, personenbezogene Profilbildung, Verkettung mit Negativmerkmalen => kein Privileg

Einfluss von PETs

- Privacy enhancing Technologies (PETs) sind zu berücksichtigen
- Art. 14 (5) (b) DSGVO „geeignete Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen betroffener Personen“
- Art. 25 (1) DSGVO Privacy by Design
- Art. 25 (2) DSGVO Privacy by Default
- Art. 32 DSGVO Sicherheit der Verarbeitung

⇒ Sind PETs vorhanden, sind sie unter Berücksichtigung des Standes der Technik auch einzusetzen!

Beispiel
Widerruf per automatisiertem Verfahren

- Art. 21 (5) DSGVO ist Ausfluss des Privacy by Design (PbD)-Prinzips: [1]

  Im Zusammenhang mit der Nutzung von Diensten der Informationsgesellschaft kann die betroffene Person ungeachtet der Richtlinie 2002/58/EG ihr Widerspruchsrecht mittels automatisierter Verfahren ausüben, bei denen technische Spezifikationen verwendet werden.

- Widerspruch soll unkompliziert sein. Denkbar automatisierte Verfahren auf Basis lokal gesetzter Voreinstellungen.[2]

[1] Paal/Pauly/Martini Art. 25 DSGVO Rn. 32.
[2] Paal/Pauly/Martini Art. 21 DSGVO Rn. 72.

Beispiel
Widerruf per automatisiertem Verfahren

- Umsetzung eines automatisierten Widerspruchs bedarf geeigneter Spezifikationen für die Kommunikation und eine geeignete Semantik zur Definition der Voreinstellungen „Privacy Preferences".

- Hier sollte auf bestehende Ergebnisse aus der Datenschutz-Forschung aufgebaut werden (P3P, PrimeLife Policy Language)

- Das Projekt „Scalable Policy-awarE linked data arChitecture for prIvacy, trAnsparency and compliance" (SPECIAL) nimmt sich Teilen dieser Aufgabe im Themenkreis Big Data an.
  Partner u.a. W3C (ERCIM), WU Wien, ULD und Unternehmen.

- https://www.specialprivacy.eu/

Lessons learned im heutigen Workshop

- Nutzen der Spezifikation aus SPECIAL für Daten aus öffentlichen Quellen, d.h. ohne vorherigen Kundenbezug. Identifier nötig? Wie vermitteln robots.txt?

- Kurator als Mittelsperson bei mehrfacher Weitergabe pseudonymisierter Daten => könnte Semantik und Beschreibung Teil einer Spezifikation werden?

Zeit für Fragen und Diskussion

Kontakt:
Benjamin Bremert
uld69@datenschutzzentrum.de
Harald Zwingelberg
uld6@datenschutzzentrum.de
www.datenschutzzentrum.de
0431/988-1222

ULD
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein