

Anonymisierung von Gerichtsurteilen

– Eine wesentliche Voraussetzung für E-Justice –

(Prof. Dr. Axel Adrian / Prof. Dr. Stefan Evert / Michael Keuchen / Philipp Heinrich / Natalie Dykes)

Ass. jur. Michael Keuchen, michael.keuchen@fau.de



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

FACHBEREICH
RECHTSWISSENSCHAFT



LEHRSTUHL FÜR
KORPUS- UND
COMPUTERLINGUISTIK

Warum Anonymisierung von Gerichtsurteilen?

Diese Frage hat mehrere Dimensionen:

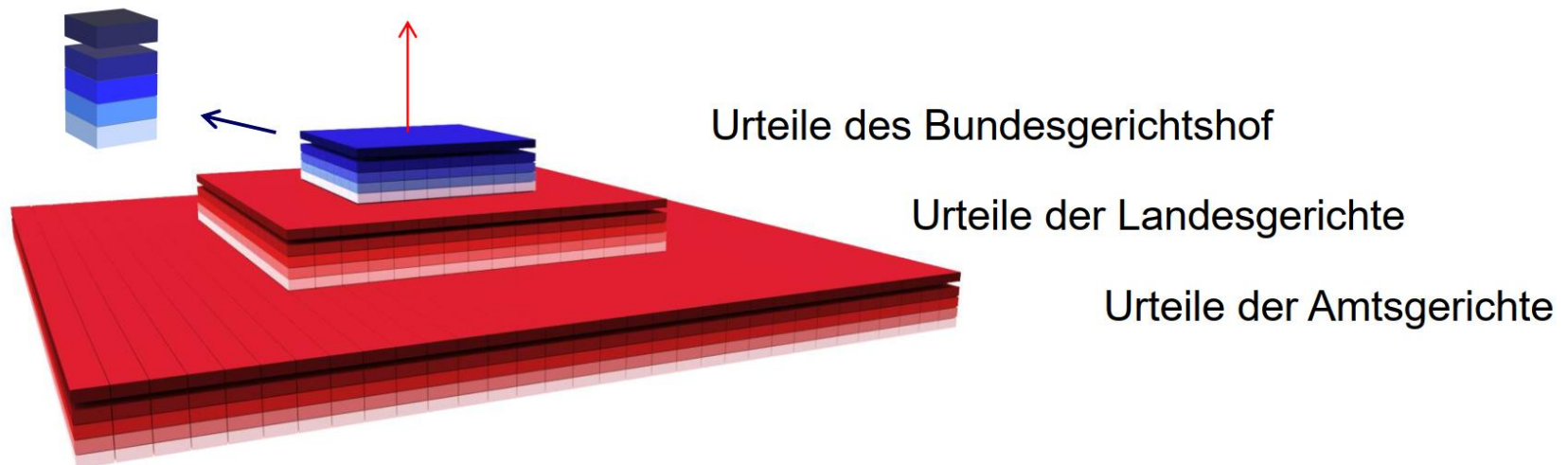
Schutz von personenbezogenen Informationen und berechtigten Interessen

Veröffentlichung von Urteilen

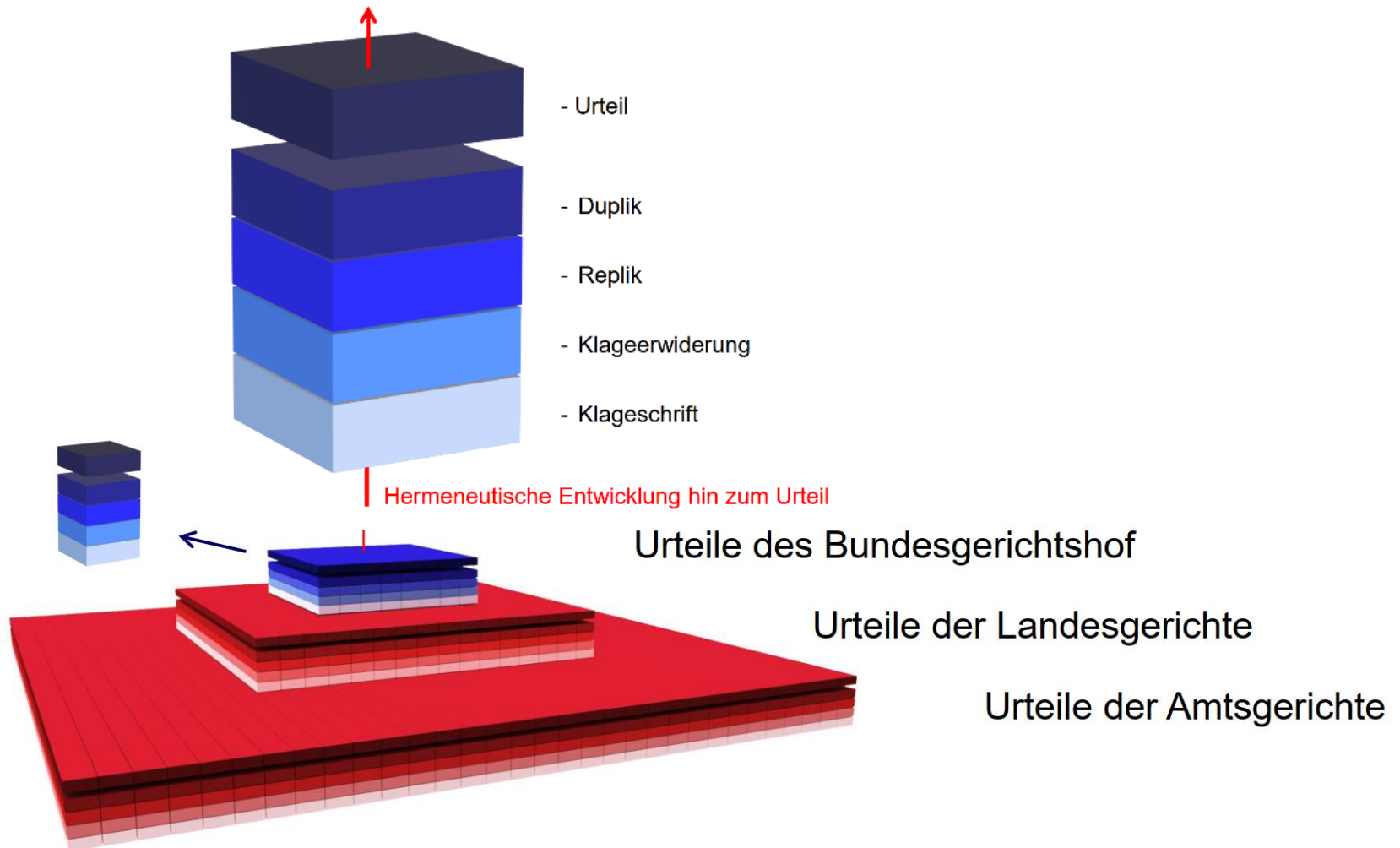
Urteile als Trainingsdaten für Legal Tech Anwendungen

Veröffentlichungspraxis in Deutschland

- Nur etwa 2 % aller Gerichtsentscheidungen werden veröffentlicht.
- Davon überwiegend letztinstanzliche und obergerichtliche Entscheidungen.
- Tatsacheninstanzen und Eingangsinstanzen sind unterrepräsentiert.



Veröffentlichungspraxis in Deutschland



Spezielle Veröffentlichungspflichten von Urteilen

In Deutschland existieren nur vereinzelt kodifizierte Veröffentlichungspflichten:

- Veröffentlichung von Entscheidungsformeln des Bundesverfassungsgerichts nach § 31 II 3 BVerfGG im Bundesgesetzblatt, soweit ein Gesetz als mit dem Grundgesetz oder sonstigem Bundesrecht vereinbar oder unvereinbar oder für nichtig erklärt wird.
- Veröffentlichung von Entscheidungsformeln des Oberverwaltungsgerichts nach § 47 V 2 HS 2 VwGO.
- Bekanntmachungen im zivilrechtlichen Musterfeststellungsverfahren, wie nach § 612 I ZPO des Musterfeststellungsurteil im Klageregister.
- ...

→ Keine kodifizierte allgemeine Veröffentlichungspflicht

Allgemeine Veröffentlichungspflicht von Urteilen

Die Rechtsprechung hat eine allgemeine Veröffentlichungspflicht entwickelt:

- Allgemeine und verfassungsimmanente Veröffentlichungspflicht aus dem Demokratie- und Rechtsstaatsgebot sowie dem Justizgewährungsanspruch.
(BVerwG NJW 1997, 2694, 2695; BGH NJW 2017, 1819, 1820)
- Publikationspflicht besteht unabhängig von der Gerichtsinstanz, Rechtskraft und eines objektiven Interesses an der Entscheidung.
- Jedoch ist eine Anonymisierung erforderlich. Der Schutz von überwiegenden Interessen der Beteiligten muss gewährleistet sein.

Was bedeutet Anonymität?

- **Anonymität** wurde durch das BVerfG u.a. im Volkszählungsurteil aus dem Recht auf informationelle Selbstbestimmung (Art. 2 II i.V.m. 1 I GG) entwickelt (BVerfG NJW 1984, 419, 422).
- *Jedem steht es frei, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen und vor unbegrenzter Erhebung, Speicherung, Verwendung und Weitergabe geschützt zu sein.*
- Die Wiege des modernen Datenschutzrechts liegt im Statistikrecht.
- **Anonymisierung:** Unsichere Verknüpfbarkeit von Merkmalsausprägungen mit einem Individuum durch Vergrößerung der Menge an möglichen Merkmalsträger.

Was bedeutet Anonymität i.S.d DSGVO?

- Betrachtet man ErwGr. 26 DSGVO, so ist entscheidend, ob sich die Information auf eine identifizierte oder identifizierbare natürliche Person bezieht.
- Für die indirekte Identifizierbarkeit, sind alle objektiven Faktoren wie die Kosten der Re-Identifizierung und der dafür erforderliche Zeitaufwand zu berücksichtigen, wobei dafür die zum Zeitpunkt der Anonymisierung verfügbare Technologie und die technologischen Entwicklungen maßgeblich sind.
- Eine faktische Anonymität ist gegeben, wenn eine Zuordnung der Einzelangaben zu einer betroffenen Person nur mit einem unvernünftig großen Aufwand an Zeit, Kosten und Arbeitskraft möglich ist.

Welche Merkmale sind zu anonymisieren?

Direkte Identifikatoren:

- Namen (natürliche und juristische Personen)
- Adressangaben
- Geburtsdaten
- ...

Indirekten Identifikatoren:

- Berufsangaben
- Titel
- Gesundheitsdaten
- deskriptive Angaben (örtliche Verhältnisse, Betriebsinformationen)
- einzigartige Merkmale (einzige rote Haus in Dechsendorf)
- ...

→ Gefahr des Cross-Referencing

Rethinking Anonymity

1. Jur. Personen sind nicht umfasst (ErwG. 14 S. 2 DSGVO).
2. DSGVO hat Individualrechtsgüterschutz im Fokus.
3. Die Justizstatistiken unterliegen bereits dem strengen Statistikrecht. Deshalb bedarf es erst Recht eines sehr strengen Schutzniveaus, wenn individuelle Datensätze in einer Vielzahl veröffentlicht werden.
4. Gefahren einer Deanonymisierung sind noch unzureichend erforscht. Empirische Daten zum Arbeits- und Zeiteinsatz sowie zum verfügbaren Zusatzwissen notwendig.

Ergebnis:

Datenschutzrecht i.S.d. DSGVO unzureichend als Maßstab für die Anonymität.

Forschungsprojekt zur automatischen Anonymisierung von Gerichtsurteilen

Ziel:

Evaluation der rechtlichen und technischen Fragen zur Möglichkeit einer automatischen Anonymisierung von Urteilen, insbesondere mit Hilfe korpuslinguistischer Verfahren.

Beteiligte:

- FAU Fachbereich Rechtswissenschaft (Prof. Dr. Adrian / Michael Keuchen) sowie FAU Lehrstuhl für Korpus- und Computerlinguistik (Prof. Dr. Evert / Philipp Heinrich / Natalie Dykes) und 4 – 8 stud. Hilfskräfte.
 - (vertretene wiss. Disziplin: Rechtsdogmatik, -theorie und Methodenlehre, Linguistik, Mathematik, Computerwissenschaft)
- Bayerisches Staatsministerium der Justiz

Datengrundlage des Projekts

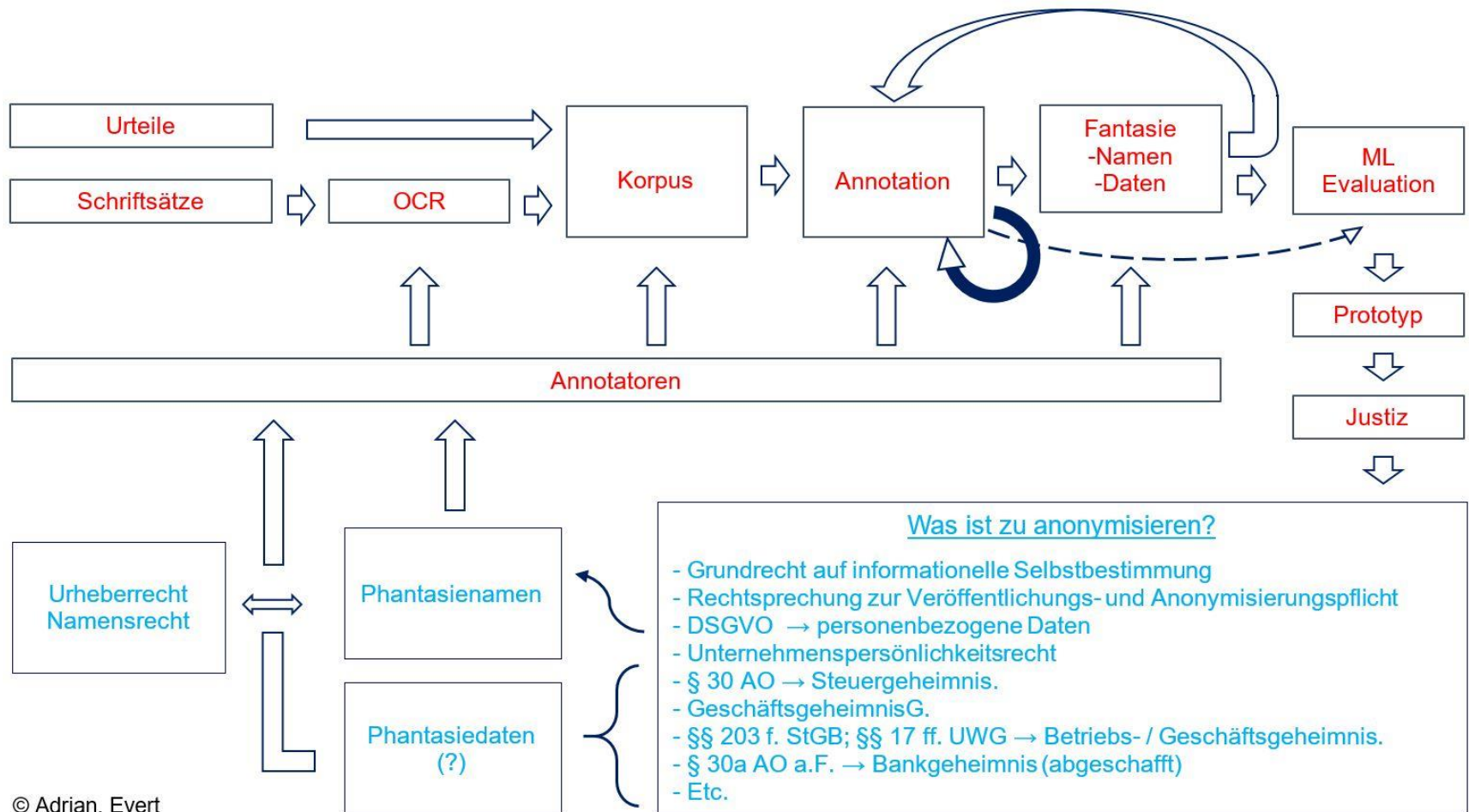
Urteile

- 604 amtsgerichtliche Urteile aus dem Verkehrsunfallrecht sowie Wohnraummietrecht in maschinenlesbarer Form.
- 1,3 Mio. Tokens (Wörter und Satzzeichen) → ca. 2.200 Tokens pro Urteil.

Dazugehörige Akten mit Schriftsätzen

- 557 Akten sind verfügbar.
- Scan der Klage, Klageerwiderung und weiteren Schriftsätzen in der Sache.
- Anwendung von Texterkennungssoftware für Maschinenlesbarkeit.

Projektablauf und Workflow



Beispiel: Annotation und Tagging

38 Ort 90762 Fürth , Gz.: 557/01

40 wegen Räumung

42 Ort erlässt das Amtsgericht Hof durch Richter am Sonstiges
Name

44 ENDURTEIL

46 I. Die Beklagten werden verurteilt, das Anwes

48 II. Die Kosten des Rechtsstreits werden gegen

50 III. Das Urteil ist vorläufig vollstreckbar. Die B

52 Tatbestand:

54 Die Kläger begehren von den Beklagten die Räumung

56 Kennzeichen Die Kläger sind Eigentümer des Anwesens Dörfelstr. 10
Ort Mietvertrag (BI. 6/16 d. A.) enthält u. a. folgende

58 Inhalt:

60 „...“

62 § 2 Nr. 2 Vertragsdauer

New Annotation

Text
Beklagten

Search
google, wikipedia

Entity type
☒ Name
☐ Telefonnummer
☐ Titel
☐ Adresse
☐ Ort
☐ Datum
☐ Kennzeichen
☐ ...

Entity attributes
☐ unsicher Risiko: ?

Notes

OK Cancel

Zwischenergebnisse zur Merkmalshäufigkeit

Nach den Annotationsrichtlinien getaggte Merkmale pro Urteil aus dem annotierten und adjudizierten Teilkorpus (300 Wohnraummietrechtsurteile).

Merkmalskategorien	Häufigkeit im Durchschnitt pro Urteil	Beispiele
Indirekt identifizierendes Merkmal einer Adressangabe	10	Rotes Haus, 235 Behandlungsplätze
Name einer natürlichen Person	10	Michael Keuchen
Adressangabe	8	Schillerstraße 1, 91054 Erlangen
Name eines juristischen Funktionsträgers	4	Ri'inAG Michaela Maier
Name einer juristischen Person	3	Mustermann AG
Aktenzeichen	3	12 C 123/21
Indirekt identifizierendes Merkmal einer nat. oder jur. Person	1	Berufe, Krankheiten, Betriebsgröße

Entwicklung von maschinellen Verfahren (Logiken)

Mensch produziert Goldstandard anhand diesem maschinelle Verfahren entwickelt, trainiert und evaluiert werden können.

Menu

Corpus queries

Standard query

Restricted query

Word lookup

Frequency lists

Keywords

Analyse corpus

Saved query data

Query history

Saved queries

Categorised queries

Bavarian Court Decisions (05/2020): powered by CQPweb

Standard Query

[Herr,Frau,Dr?,Prof?,Zeuge,Richter*] {????*/N}

Query mode: [Simple query language syntax](#)

Number of hits per page:

Match strategy:

LOGIK

44	<p>t2018_01 Adrian 00_Z_BECKRS_B_2018 Adrian 5</p>	<p>Morgenpost "angemerkt hat". Ein weiteres Beispiel für die Erosion von</p> <p>6000 Euro übersteigender Abfindungsbetrag mit an Sicherheit grenzender Wahrscheinlichkeit nicht zusätzlich zur Verfügung stehen würde. Auch hier haben sich die Klägerin und der Beklagte konträr eingelassen und auch der Zeuge Adrian hat die Angaben der Klägerin bestätigt. Auch hier kann ein Beweis mit dem Maßstab des § 286 ZPO nicht geführt werden. Insoweit wird auf die Bewertung der Einlassungen</p>
----	--	---

Wir erstellen durch Annotation des zu Anonymisierenden Trainingsdaten.
Die Maschine sucht mit deren Hilfe (umgekehrt zum obigen Beispiel) Muster in der Sprache, um das zu Anonymisierende zu finden. So könnte es gelingen, dass die Maschine automatisch anonymisiert.

Ausblick und angedachte Ergebnisse des Projekts

- Juristische Anonymisierungsdogmatik
- Goldstandard für die Anonymisierung von Urteilen
- Hermeneutische Betrachtung mit Hilfe der Schriftsätze
- Umfassender anonymer Datensatz aus Urteilen und Schriftsätzen
(nach Freigabe durch das Bayerische Staatsministerium der Justiz)



Vielen Dank für Ihr Interesse!

Ass. jur. Michael Keuchen, michael.keuchen@fau.de



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

FACHBEREICH
RECHTSWISSENSCHAFT



LEHRSTUHL FÜR
KORPUS- UND
COMPUTERLINGUISTIK