

Kurz-Präsentation Entscheidungsanonymisierung

Projektüberblick

September 2022

Dr. Martin Schneider

Abteilung für Rechtsinformatik, Informations-
und Kommunikationstechnologie

Anonymisierung von Gerichtsentscheidungen: Status Quo



Publizierung von Gerichtsentscheidungen

- Gerichtsentscheidungen werden im **Rechtsinformationssystem (RIS)** der Republik Österreich veröffentlicht, das unter www.ris.bka.gv.at aufgerufen werden kann
- Eine **Verpflichtung zur Veröffentlichung** besteht nur für den **Obersten Gerichtshof** (§ 15 OGH-G). Ebendort sind auch die Vorgaben für die Anonymisierung festgelegt.
- Dies hat zur Folge, dass fast ausschließlich Entscheidungen des Obersten Gerichtshof publiziert sind. Die **Entscheidungen anderer Gerichte** werden intern und nicht anonymisiert aufbewahrt.
- Jeder spricht über OpenJustice, Access2Justice, Transparenz, etc. aber warum sollten wir nicht mehr Entscheidungen veröffentlichen? => **Regierungsprogramm**



Herausforderungen

- Die Anonymisierung von Gerichtsentscheidungen ist ein **teil-manueller Prozess**, der zwar durch spezifische Funktionen in LibreOffice unterstützt wird, aber trotzdem viel Zeit benötigt
- Unzureichende Personalkapazitäten an Gerichten



Vorgehensweise

- Einsatz mehrerer Technologien aus dem Bereich der **künstlichen Intelligenz zur Automatisierung** der Anonymisierungstätigkeiten

Grundlagen aktuell

- Oberster Gerichtshof, § 14 OGHG, BGBl. Nr. 328/1968 **seit 1991!**
 - alle Entscheidungen veröffentlichen (Ausnahme: rein formale E)
 - anonymisieren, soweit Verständlichkeit gewahrt bleibt
 - wenn Anonymität nicht gesichert ist, nicht veröffentlichen
 - Gericht entscheidet diese Fragen
- Gerichtsbarkeit außer OGH § 48a GerichtsorganisationsG, RGBl. Nr. 217/1896
 - Rechtskräftige E anderer Gerichte von grundsätzlicher Bedeutung
 - Bestimmungen d OGHG anzuwenden
- Regierungsprogramm 2020 -2024
 - „Verpflichtende Veröffentlichung im RIS zumindest von Urteilen der Oberlandesgerichte“

Fragen

- Transparenz oder Rechtsinformation?
- Alle Entscheidungen – ausgewählte?
 - Alle Instanzen ?
 - Nur rechtskräftige?
 - Nur rechtlich bedeutende?
- Anonymisieren?
- Gerichtsbarkeit oder Justizverwaltung?
- Wer entscheidet über Veröffentlichung?
 - Gericht – Justizdienststelle – externer Verarbeiter?
- Wer entscheidet über Anonymisierung?
 - Gericht – Justizdienststelle – externer Verarbeiter?

Situation in Österreich

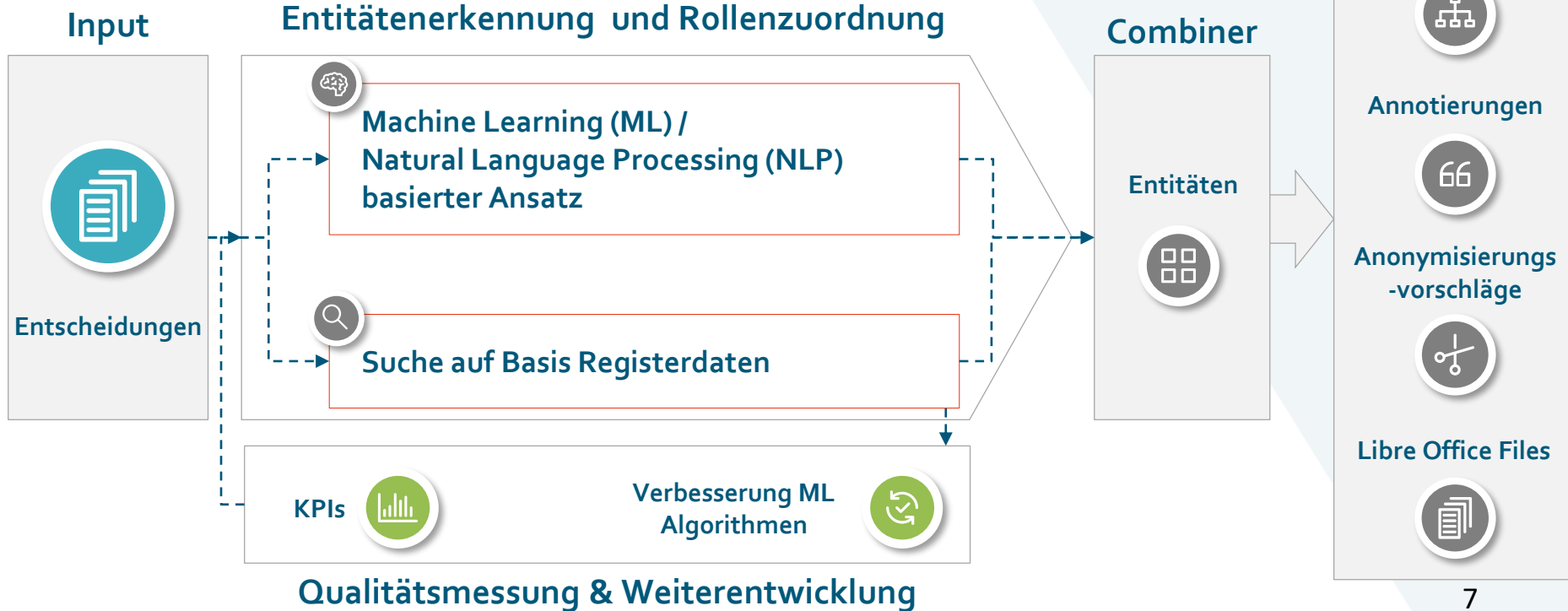
(außer OGH)

- Früher
 - Entscheidungssammlungen der Gerichte nicht anonymisiert
 - justizintern
 - uneinheitlich
 - Kaum Veröffentlichungen und willkürliche Auswahl
- Jetzt
 - alles anonymisiert
 - bundesweiter Zugriff derzeit nur für Gerichte => Optimierung
 - Veröffentlichung ausgewählter OLG Entscheidungen => menschliche Nachkontrolle!

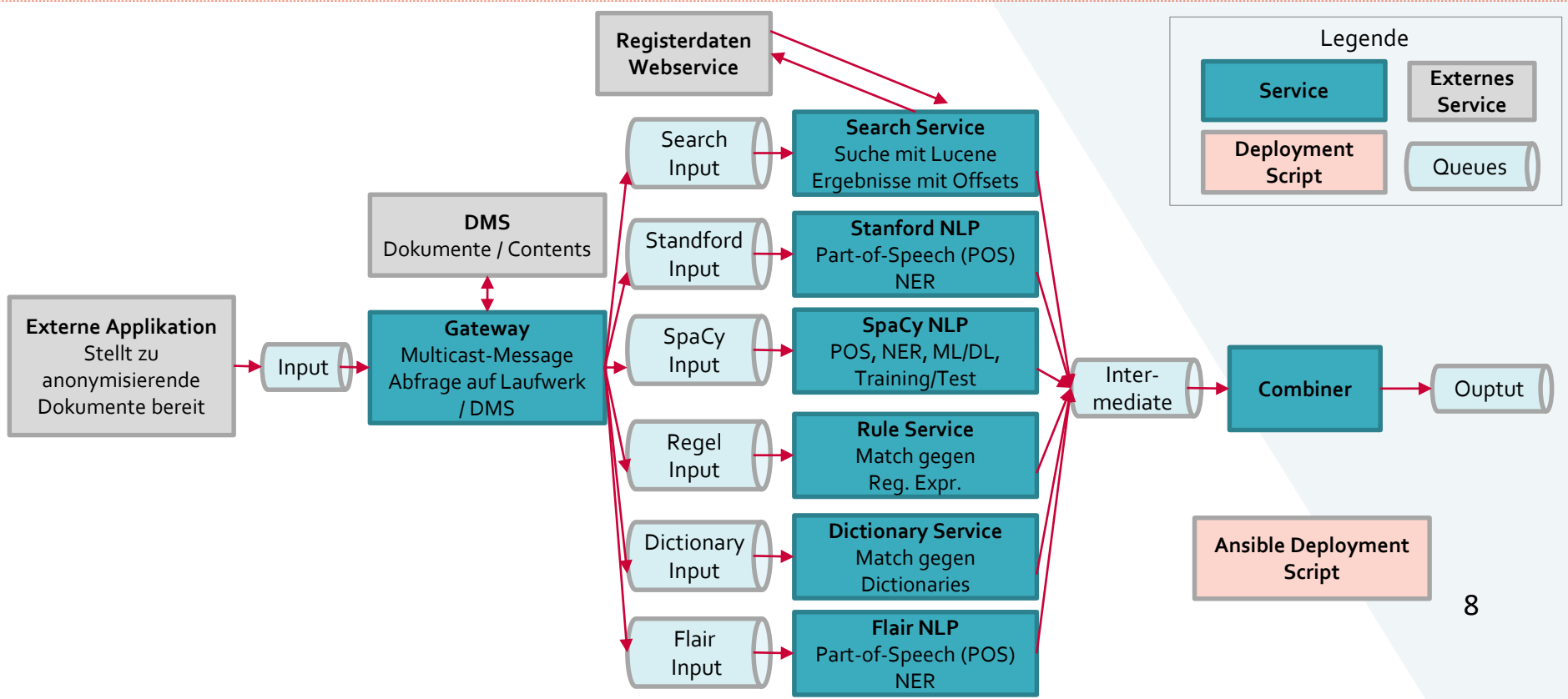
Situation in Österreich: künftig

- Überlegungen zur Neugestaltung der Veröffentlichung
 - OLG Entscheidungen: rechtskräftig und grundsätzlich
 - auch Vorentscheidungen, wenn zum Verständnis nötig
 - Auswahl und Anonymisierung
 - durch Justizverwaltung
 - => Gleichmäßigkeit und mehr Veröffentlichungen
 - Anhörung des Gerichts
 - Clearingstelle beim OGH vor Veröffentlichung
 - Anonymisierung aus DJAP angestoßen, Workflow für Veröffentlichung
 - Vorerst keine automatische Veröffentlichung

High-Level Architektur



Architektur Ebene 1



Bestehende Datensets (für Training / ML)



- Entscheidungen des Obersten Gerichtshofs
– Geschwärzte Dokumente mit hinterlegtem Original-Text
– Registerdaten mit Informationen zu Verfahrensbeteiligten

ca. 66.000

- Entscheidungen „anderer“ Gerichte
– Nicht annotierte Dokumente mit Original-Text
– Registerdaten mit Informationen zu Verfahrensbeteiligten

ca. 800.000

- Ergänzende Daten
– Öffentlich verfügbare Verzeichnisse (z.B. Adressen, Literatur ...)

Anonymisierungsregeln

- Folgende anonymisierungsrelevanten Informationen werden mit einem **fortlaufenden Buchstaben und einem Stern (*) anonymisiert** (z.B. beginnend mit A*), wobei ein einmal für die konkrete Information vergebener Buchstaben in der gesamten Entscheidung für diese beibehalten wird:
 - Personen
 - Richter*innen im Rahmen eines Ablehnungsverfahrens
 - Anonymisierungsrelevante Informationen, die mehr als einmal im Dokument vorkommen

Anonymisierungsregeln

- Darüber hinaus werden folgende nur einmal im Dokument vorkommende Informationen bzw. Daten abweichend mit **zwei Sternen (**)** versehen:
 - Produktbezeichnungen
 - Wohnort, Sitz und Adressen
 - Kontodaten
 - Geburts- und Sterbedaten
 - E-Mailadressen & Internetadressen
 - Firmenbuchnummern
 - Katastralgemeinden und Einlagezahlen

Anonymisierungsregeln

- Basierend auf den fachlichen Vorgaben werden unter anderem folgende, teils personenbezogene, Informationen in der Regel **nicht anonymisiert**
 - Richter*innen, Laienrichter*innen sowie Schriftführer*innen (ausgenommen im Rahmen von Ablehnungsverfahren)
 - Ausgewählte öffentliche Institutionen (Gerichte, Staatsanwaltschaften, Bundesministerien)
 - Berufsmäßige Parteienvertreter:innen (inkl. Anschrift)
 - Gebietsbezeichnungen (z.B. Bundesländer, Länder, Kontinente)
 - Gesetze, (jur.) Fachliteratur, Stichtage etc.
- Darüber hinaus werden akademische Titel, Namenszusätze wie Junior oder Senior sowie Rechtsformzusätze nicht anonymisiert,

Output – Dokumenten Vergleich

Original

Der Oberste Gerichtshof hat durch den Senatspräsidenten Dr. Hoch als Vorsitzenden in der Rechtssache der klagenden Partei Max Mustermann, vertreten durch ABC Rechtsanwälte OG in Teststadt, wider die beklagte Partei Michaela Musterfrau, vertreten durch Mag. Müller, Rechtsanwalt in Musterdorf, wegen Einwendungen gegen den Anspruch (§ 35 EO), in nichtöffentlicher Sitzung zu Recht erkannt:

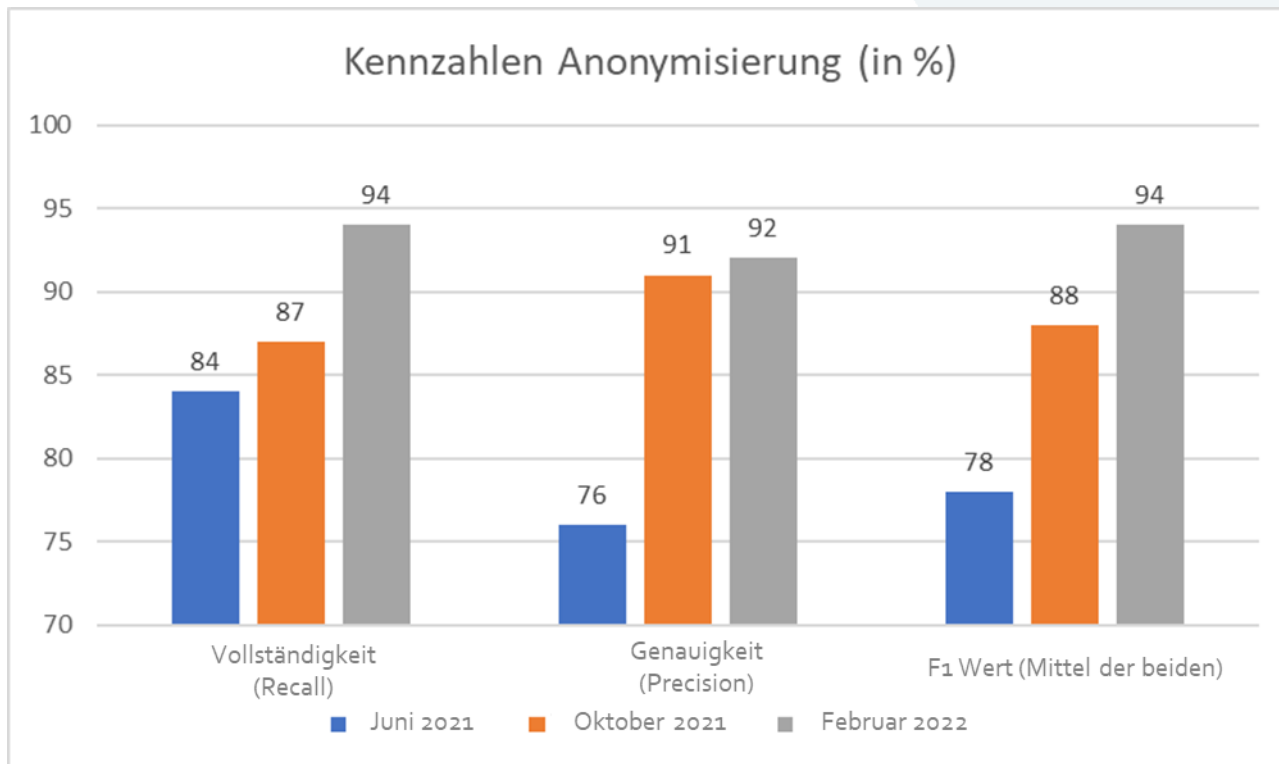
Annotiert

Der Oberste Gerichtshof hat als Revisionsgericht durch den Senatspräsidenten Dr. Hoch als Vorsitzenden sowie die Hofräte Dr. Roch und Priv.-Doz. Dr. Rassi in der Rechtssache der klagenden Partei ~~Max Mustermann A*~~, wohnhaft in ~~Wiener Neustadt **~~, vertreten durch ABC Rechtsanwälte OG in Teststadt, wider die beklagte Partei ~~Michaela Musterfrau B*~~ vertreten durch Mag. Müller, Rechtsanwalt in Musterdorf, wegen Einwendungen gegen den Anspruch (§ 35 EO), in nichtöffentlicher Sitzung zu Recht erkannt:

Anonymisiert

Der Oberste Gerichtshof hat als Revisionsgericht durch den Senatspräsidenten Dr. Hoch als Vorsitzenden sowie die Hofräte Dr. Roch und Priv.-Doz. Dr. Rassi in der Rechtssache der klagenden Partei A*, wohnhaft in **, vertreten durch ABC Rechtsanwälte OG in Teststadt, wider die beklagte Partei B* vertreten durch Mag. Müller, Rechtsanwalt in Musterdorf, wegen Einwendungen gegen den Anspruch (§ 35 EO), in nichtöffentlicher Sitzung zu Recht erkannt:

KPIs Entscheidungsanonymisierung



Zeitverlauf Anonymisierungsprojekt



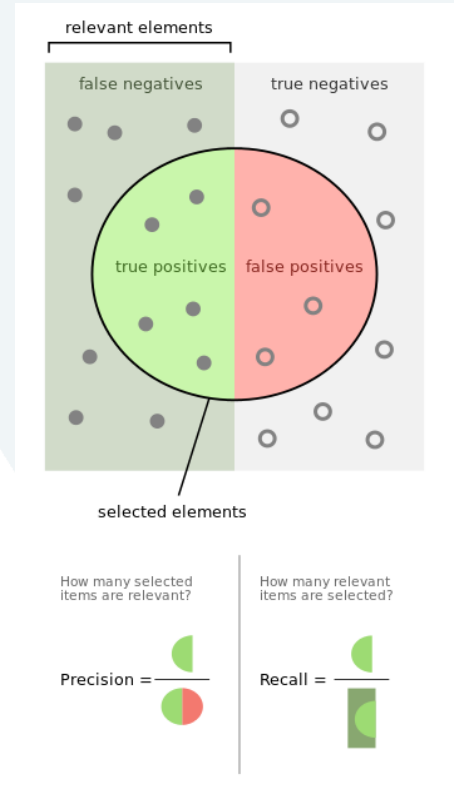
Vielen Dank!

Recall

recall / completeness:

Wieviele der relevanten Elemente wurden ausgewählt?

- Maximieren, wenn möglichst viele der zu findenden Elemente gefunden werden sollen.
- Maximierung steigert Anzahl der (falsch) gefunden Elemente



Ground Truth

Die Zeugin Klara identifizierte die Täter Alex und Daniel im Supermarkt.

High recall

Die Zeugin Klara identifizierte die Täter Alex und Daniel im Supermarkt.

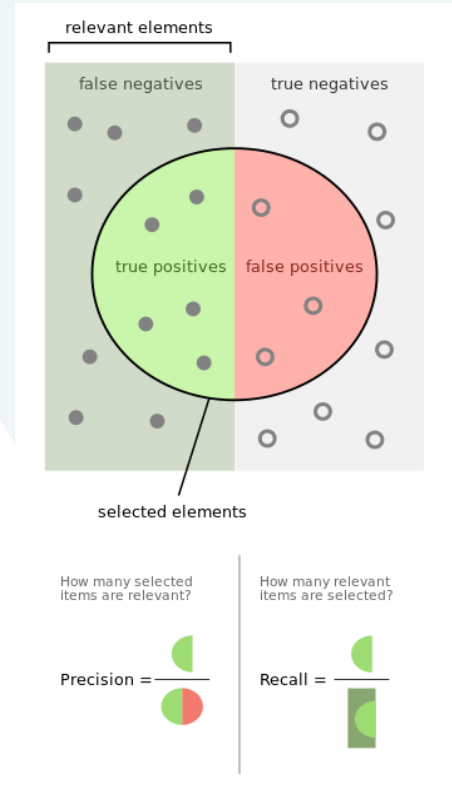
Recall: 100%; Precision: 75%

Precision

precision / correctness:

Welche der erkannten Elemente sind relevant?

- Maximieren, wenn man sich sicher sein will, dass nur „richtige“ Elemente ausgewählt wurden.
- Maximierung steigert Korrektheit der Resultate, wird aber viele Elemente auslassen.



Ground
Truth

Die Zeugin Klara identifizierte die Täter Alex und Daniel im Supermarkt.

High
precision

Die Zeugin Klara identifizierte die Täter Alex und Daniel im Supermarkt.

Recall: 66%; Precision: 100%

f1_score, f2_score

precision / correctness:

Welche der erkannten Elemente sind relevant?

recall / completeness:

Wieviele der relevanten Elemente wurden ausgewählt?

f1_score *:

Verhältnis zwischen precision & recall

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$$

- Höher bedeutet korrekter und gleichzeitig mehr Elemente gefunden

f2_score *:

Verhältnis zwischen precision & recall, höhere Gewichtung der vollständigen Abdeckung gegenüber vollständiger Korrektheit.

- Gewichtungsfaktor (2) erhöht gegenüber (1)

