Corpus der Entscheidungen des Bundespatentgerichts

(CE-BPatG)

Codebook

Version 2024-07-09

DOI: 10.5281/zenodo.10849977

Titel	Corpus der Entscheidungen des Bundespatentgerichts
Abkürzung	CE-BPatG
Autor	Seán Fobbe
Version	2024-07-09
Download	https://doi.org/10.5281/zenodo.10849977
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Seán Fobbe (2024). Corpus der Entscheidungen des Bundespatentgerichts (CE-BPatG). Version 2024-07-09. Zenodo. DOI: 10.5281/zenodo.10849977.

Digital Object Identifier (DOI): Concept DOI und Version DOI

Soweit nicht anders angegeben ist die DOI immer eine »Version DOI« und bezieht sich nur auf eine bestimmte Version des Datensatzes. Sie verweist daher nur auf Version 2024-07-09. Für das Gesamtkonzept dieses Datensatzes steht eine »Concept DOI« zur Verfügung, die auf der Zenodo-Seite jeder Version unter »Cite all versions?« zu finden ist. Sie lautet 10.5281/zenodo.3954850. Die »Concept DOI« verlinkt immer die aktuellste Version.

Urheberrecht

Disclaimer

Dieser Datensatz ist eine private wissenschaftliche Initiative und steht in keiner Verbindung zu Behörden, Gerichten oder anderen amtlichen Stellen der Bundesrepublik Deutschland.

Inhaltsverzeichnis

1 Einführung

Das **Bundespatentgericht (BPatG)** ist ein am 1. Juli 1961 speziell für den gewerblichen Rechtsschutz (z.b. Patente und Marken) errichtetes oberes Bundesgericht (Art. 96 Abs. 1 GG). Es ist für die Kontrolle von Entscheidungen des Deutschen Patent- und Markenamts und des Bundessortenamts, sowie Nichtigkeitsklagen gegen Patente und für Zwangslizenzverfahren zuständig (§§ 65 Abs. 1 PatG). Im Instanzenzug der ordentlichen Gerichtsbarkeit ist es nur dem Bundesgerichtshof untergeordnet.[1] Es hat seinen Sitz in München am Sitz des Deutschen Patent- und Markenamts (§ 65 Abs. 1 PatG).

Im Jahr 2022 besteht des BPatG aus 24 Senaten: ein Juristischer Beschwerdesenat, 6 Nichtigkeitssenate, 10 Technische Beschwerdesenate, 4 Marken-Beschwerdesenate, ein Marken- und Design-Beschwerdesenat, ein Gebrauchsmuster-Beschwerdesenat und ein Beschwerdesenat für Sortenschutzsachen. Die Besetzung der Senate unterscheidet sich je nach Verfahrensart und beträgt zwischen drei und fünf Richter:innen (§ 67 BPatG).

Die Richterschaft des Bundespatengerichts unterscheidet sich merklich von der anderer Gerichte: in allen Verfahren in denen technischen Erfindungen zu beurteilen sind müssen sowohl Jurist:innen als auch »technische Richter:innen« mitwirken. Technische Richter:innen müssen ein technisches oder naturwissenschaftliches Studium, 5 Jahre Berufserfahrung und entsprechende Rechtskenntnisse vorweisen können (§§ 65 Abs. 2 S. 3, 26 Abs. 3 PatG). Nur markenrechtliche Verfahren werden von Jurist:innen alleine geführt. Insgesamt 99 Richter:innen sind aktuell am Bundespatentgericht tätig, davon 57 technische Richter:innen (Stichtag 31. Dezember 2021).

Wieso dieser Datensatz? Die quantitative Analyse von juristischen Texten, insbesondere denen des BPatG, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.[2] Zu einem nicht unerheblichen Teil liegt dies auch daran, dass die Anzahl an frei nutzbaren Datensätzen außerordentlich gering ist.

Die meisten hochwertigen Datensätze lagern (fast) unerreichbar in kommerziellen Datenbanken und sind wissenschaftlich gar nicht oder nur gegen Entgelt zu nutzen. Frei verfügbare Datenbanken wie *Opinio Iuris*[3] und *openJur*[4] verbieten ausdrücklich das maschinelle Auslesen der Rohdaten. Wissenschaftliche Initiativen wie der Juristische Referenzkorpus (JuReKo) sind nach jahrelanger Arbeit hinter verschlossenen Türen verschwunden.

In einem funktionierenden Rechtsstaat muss die Rechtsprechung öffentlich, transparent und nachvollziehbar sein. Im 21. Jahrhundert bedeutet dies auch, dass sie systematischer Überprüfung mittels quantitativen Analysen zugänglich sein muss. Der Erstellung und Aufbereitung des Datensatzes liegen daher die Prinzipien der allgemeinen Verfügbarkeit durch Urheberrechtsfreiheit, strenge Transparenz und vollständige wissenschaftliche Reproduzierbarkeit zugrunde. Die FAIR-Prinzipien (Findable, Accessible, Interoperable and Reusable) für freie wissenschaftliche Daten inspirieren sowohl die Konstruktion, als auch die Art der Publikation.[5]

---

[1] Die »ordentliche Gerichtsbarkeit« ist eine historische gewachsene Bezeichnung. Früher war die Verwaltungsgerichtsbarkeit nicht mit unabhängigen Richtern, sondern mit Verwaltungsbeamten besetzt und daher »außerordentlich«. Die mit unabhängigen Richtern besetzten Gerichte wurden als »ordentlich« bezeichnet.

[2] Besonders positive Ausnahmen finden sich unter: https://www.quantitative-rechtswissenschaft.de/

[3] https://opinioiuris.de/

[4] https://openjur.de/

[5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

2 Nutzung

Die Daten sind in offenen, interoperablen und weit verbreiteten Formaten (CSV, TXT, PDF) veröffentlicht. Sie lassen sich grundsätzlich mit allen modernen Programmiersprachen (z.B. Python oder R), sowie mit grafischen Programmen nutzen.

Wichtig: Nicht vorhandene Werte sind sowohl in den Dateinamen als auch in den CSV-Dateien mit »NA« codiert.

2.1 CSV-Dateien

Am einfachsten ist es die **CSV-Dateien** einzulesen. CSV[6] ist ein einfaches und maschinell gut lesbares Tabellen-Format. In diesem Datensatz sind die Werte komma-separiert. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Die Variablen sind unter Abschnitt 5 genauer erläutert.

Zum Einlesen empfehle ich für **R** dringend das package **data.table** (via CRAN verfügbar). Dessen Funktion **fread()** ist etwa zehnmal so schnell wie die normale **read.csv()**-Funktion in Base-R. Sie erkennt auch den Datentyp von Variablen sicherer. Ein Vorschlag:

```
library(data.table)
dt.bpatg <- fread("filename.csv")
```

2.2 TXT-Dateien

Die **TXT-Dateien** inklusive Metadaten können zum Beispiel mit **R** und dem package **readtext** (via CRAN verfügbar) eingelesen werden. Ein Vorschlag:

```
library(readtext)
df.bpatg <- readtext("./*.txt",
                     docvarsfrom = "filenames",
                     docvarnames =  c("gericht",
                                      "senatsgruppe",
                                      "leitsatz",
                                      "datum",
                                      "spruchkoerper_az",
                                      "registerzeichen",
                                      "eingangsnummer",
                                      "eingangsjahr_az",
                                      "zusatz_az",
                                      "kollision"),
                     dvsep = "_",
                     encoding = "UTF-8")
```

[6] Das CSV-Format ist in RFC 4180 definiert, siehe https://tools.ietf.org/html/rfc4180

3 Konstruktion

3.1 Beschreibung des Datensatzes

Dieser Datensatz ist eine digitale Zusammenstellung von möglichst allen begründeten Entscheidungen, die auf der amtlichen Internetpräsenz des Bundespatentgerichts (BPatG) am jeweiligen Stichtag veröffentlicht waren. Die Stichtage für jede Version entsprechen exakt der Versionsnummer.

Zusätzlich zu den aufbereiteten maschinenlesbaren Formaten (TXT und CSV) sind die PDF-Rohdaten enthalten, damit Analyst:innen gegebenenfalls ihre eigene Konvertierung vornehmen können. Die PDF-Rohdaten wurden inhaltlich nicht verändert und nur die Dateinamen angepasst, um die Lesbarkeit für Mensch und Maschine zu verbessern.

Speziell an Praktiker:innen richtet sich die PDF-Sammlung aller Leitsatzentscheidungen.

3.2 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	https://www.bundespatentgericht.de/
Source Code	https://doi.org/10.5281/zenodo.10849980
Registerzeichen	https://doi.org/10.5281/zenodo.4569564

Die Tabelle der Registerzeichen und der ihnen zugeordneten Verfahrensarten stammt aus dem folgenden Datensatz: »Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.«

3.3 Sammlung der Daten

Die Daten wurden unter Beachtung des Robot Exclusion Standard (RES) gesammelt. Der Abruf geschieht ausschließlich über TLS-verschlüsselte Verbindungen. Die Entscheidungen sind laut dem Gericht anonymisiert, aber ungekürzt.

3.4 Source Code und Compilation Report

Der gesamte Source Code — sowohl für die Erstellung des Datensatzes, als auch für dieses Codebook — ist öffentlich einsehbar und dauerhaft erreichbar im wissenschaftlichen Archiv des CERN unter dieser Addresse hinterlegt: https://doi.org/10.5281/zenodo.10849980

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem Codebook). Der Compilation Report enthält den vollständigen und kommentierten Source Code, dokumentiert relevante Rechenergebnisse, gibt sekundengenaue Zeitstempel an und ist mit einem klickbaren Inhaltsverzeichnis versehen. Er ist zusammen mit dem Source Code hinterlegt. Wenn Sie sich für Details der Herstellung interessieren, lesen Sie diesen bitte zuerst.

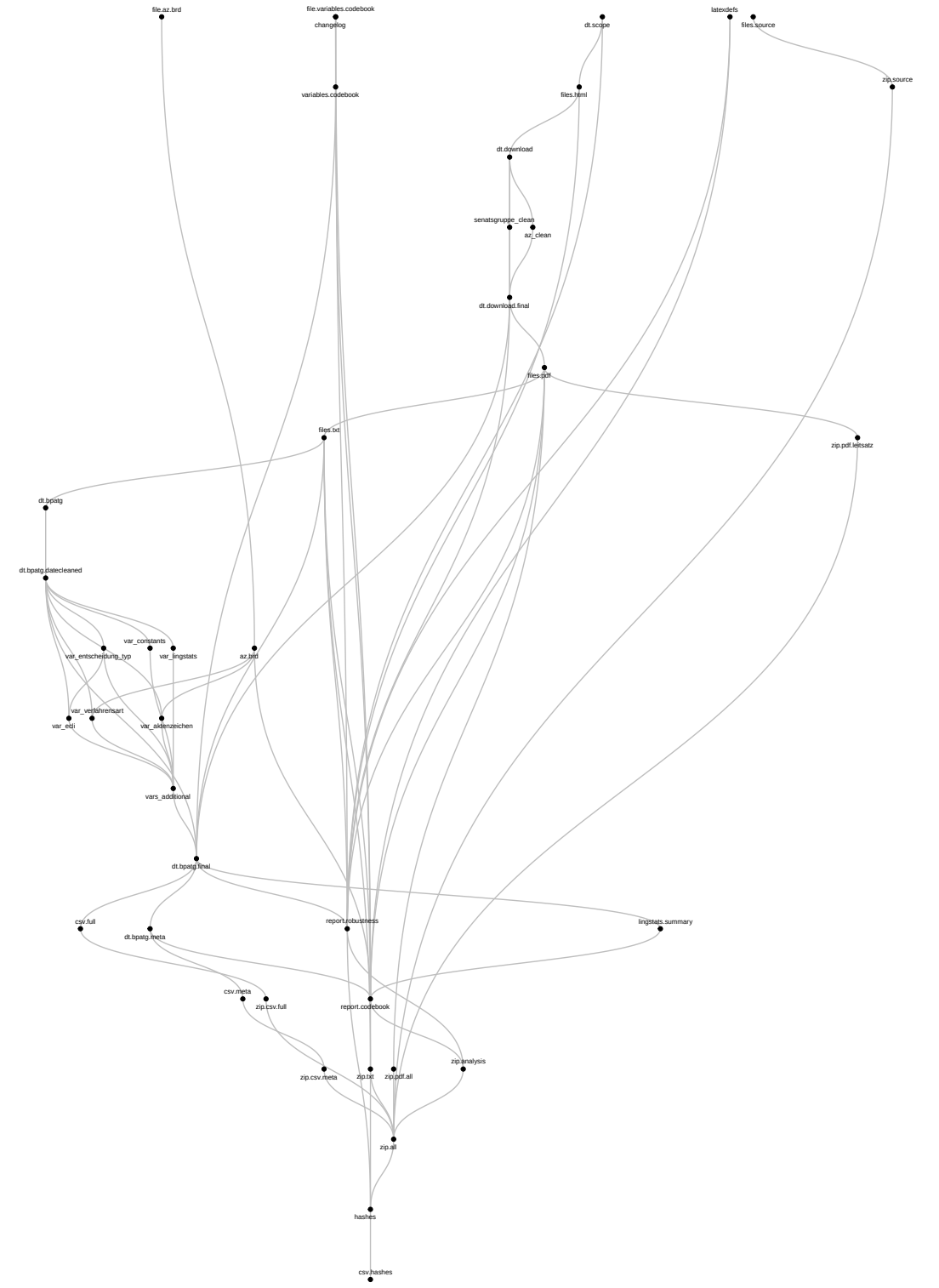


Abbildung 1: Der vollständige Prozess der Datensatz-Kompilierung.

3.5 Grenzen des Datensatzes

Nutzer:innen sollten folgende wichtige Grenzen beachten:

- Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*).
- Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Entscheidungen nicht oder nicht korrekt abgerufen werden (*automation bias*).
- Erst ab dem 1. Januar 2000 sind begründete Entscheidungen des Bundespatentgerichts einigermaßen vollständig veröffentlicht (*temporal bias*). Die Frequenztabellen geben hierzu genauer Auskunft.

3.6 Urheberrechtsfreiheit von Rohdaten und Datensatz

An den Entscheidungstexten und amtlichen Leitsätzen besteht gem. § 5 Abs. 1 UrhG kein Urheberrecht, da sie amtliche Werke sind. § 5 UrhG ist auf amtliche Datenbanken analog anzuwenden (BPatG, Beschluss vom 28.09.2006, I ZR 261/03, »Sächsischer Ausschreibungsdienst«).

Alle eigenen Beiträge (z.B. durch Zusammenstellung und Anpassung der Metadaten) und damit den gesamten Datensatz stelle ich gemäß einer *CC0 1.0 Universal Public Domain Lizenz* vollständig urheberrechtsfrei.

3.7 Metadaten

3.7.1 Allgemein

Die Metadaten in den Dateinamen sind größtenteils unverändert von den jeweiligen Datenbankeinträgen aus der amtlichen Datenbank des Bundespatentgerichts entnommen. Berechnet und hinzugefügt wurden durch den Autor des Datensatzes eine Reihe weitere Variablen, sowie in den Dateinamen der PDF/TXT-Dateien Unter- und Trennstriche, um die Maschinenlesbarkeit zu erleichtern.

Der volle Satz an Metadaten ist nur in den CSV-Dateien enthalten. Alle hinzugefügten Metadaten sind vollständig maschinenlesbar dokumentiert. Sie sind entweder im Source Code enthalten, mit dem Source Code zusammen dokumentiert oder über dauerhaft stabile Identifikatoren (z.B. DOI) zitiert.

Die Dateinamen der PDF- und TXT-Dateien enthalten Gerichtsname, Senatsgruppe, die Angabe ob es sich um eine Leitsatzentscheidung handelt, Datum, das offizielle Aktenzeichen, ggf. einen Zusatz zum Aktenzeichen und eine durch den Autor des Datensatzes generierte Kollisions-ID.

3.7.2 Schema für die Dateinamen

```
[gericht]_[senatsgruppe]_[leitsatz]_[datum]_[spruchkoerper_az]_
[registerzeichen]_[eingangsnummer]_[eingangsjahr_az]_[zusatz_az]_
[kollision]
```

3.7.3 Beispiel eines Dateinamens

```
BPatG_GebrM_LE_2011-04-14_35_W-pat_26_10_NA_0.pdf
```

3.8 Qualitätsprüfung

Die Typen der Variablen wurden mit *regular expressions* strikt validiert. Die möglichen Werte der jeweiligen Variablen wurden zudem durch Frequenztabellen und Visualisierungen auf ihre Plausibilität geprüft. Insgesamt werden zusammen mit jeder Kompilierung Dutzende Tests zur Qualitätsprüfung durchgeführt. Alle Ergebnisse der Qualitätsprüfungen sind aggregiert im Bericht »Robustness Checks« zusammen mit dem Source Code und einzeln im Archiv »ANALYSE« zusammen mit dem Datensatz veröffentlicht.

3.9 Grafische Darstellung

Die Robenfarbe der Richter des Bundespatentgerichts ist schwarz, mit stahlblauen Besätzen. Der Hex-Wert für stahlblau ist #005189. Das ist besonders bei der Erstellung thematisch passender Diagrammen hilfreich. Alle im Compilation Report und diesem Codebook präsentierten Diagramme sind in diesem stahlblau gehalten.

4 Varianten und Zielgruppen

Dieser Datensatz ist in verschiedenen Varianten verfügbar, die sich an unterschiedliche Zielgruppen richten. Zielgruppe sind nicht nur quantitativ forschende Rechtswissenschaftler:innen, sondern auch traditionell arbeitende Jurist:innen. Idealerweise müssen quantitative Methoden ohnehin immer durch qualitative Interpretation, Theoriebildung und kritische Auseinandersetzung verstärkt werden (*mixed methods approach*).

Lehrende werden von den vorbereiteten Tabellen und Diagrammen besonders profitieren, die bei der Erläuterung der Charakteristika der Daten hilfreich sein können und Zeit im universitären Alltag sparen. Alle Tabellen und Diagramme liegen auch als separate Dateien vor um sie einfach z.B. in Präsentations-Folien oder Handreichungen zu integrieren.

Variante	Zielgruppe und Beschreibung
PDF	**Traditionelle juristische Forschung**. Die PDF-Dokumente wie sie vom BPatG auf der amtlichen Webseite bereitgestellt werden, jedoch verbessert durch semantisch hochwertige Dateinamen, die der leichteren Auffindbarkeit von Entscheidungen dienen. Die Dateinamen sind so konzipiert, dass sie auch für die traditionelle qualitative juristische Arbeit einen erheblichen Mehrwert bieten. Im Vergleich zu den CSV-Dateien enthalten die Dateinamen nur einen reduzierten Umfang an Metadaten, um Kompatibilitätsprobleme zu vermeiden und die Lesbarkeit zu verbessern. Neben dem vollen Datensatz ist für Praktiker:innen auch eine Variante aufbereitet, die *Leitsatzentscheidungen* enthält.
CSV_Datensatz	**Legal Tech/Quantitative Forschung**. Diese CSV-Datei ist die für statistische Analysen empfohlene Variante des Datensatzes. Sie enthält den Volltext aller Entscheidungen, sowie alle in diesem Codebook beschriebenen Metadaten. Über Zeilenumbrüche getrennte Wörter wurden zusammengefügt.
CSV_Metadaten	**Legal Tech/Quantitative Forschung**. Wie die andere CSV-Datei, nur ohne die Entscheidungstexte. Sinnvoll für Analyst:innen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen.
TXT	**Subsidiär für alle Zielgruppen**. Diese Variante enthält die vollständigen aus den PDF-Dateien extrahierten Entscheidungstexte, aber nur einen reduzierten Umfang an Metadaten, der dem der PDF-Dateien entspricht. Die TXT-Dateien sind optisch an das Layout der PDF-Dateien angelehnt. Geeignet für qualitativ arbeitende Forscher:innen, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben oder für quantitativ arbeitende Forscher:innen, die beim Einlesen der CSV-Dateien Probleme haben. Über Zeilenumbrüche getrennte Wörter wurden *nicht* zusammengefügt.

Variante	Zielgruppe und Beschreibung
ANALYSE	**Alle Lehrenden und Forschenden**. Dieses Archiv enthält alle während dem Kompilierungs- und Prüfprozess erstellten Tabellen (CSV) und Diagramme (PDF, PNG) im Original. Sie sind inhaltsgleich mit den in diesem Codebook verwendeten Tabellen und Diagrammen. Das PDF-Format eignet sich besonders für die Verwendung in gedruckten Publikationen, das PNG-Format besonders für die Darstellung im Internet. Analyst:innen mit fortgeschrittenen Kenntnissen in R können auch auf den Source Code zurückgreifen. Empfohlen für Nutzer:innen die einzelne Inhalte aus dem Codebook für andere Zwecke (z.B. Präsentationen, eigene Publikationen) weiterverwenden möchten.

5 Variablen

5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame':   30866 obs. of  31 variables:
##  $ doc_id           : chr  "BPatG_GebrM_LE_2006-02-23_5_W-pat_429_05_NA_0.txt"
    "BPatG_GebrM_LE_2006-07-04_5_W-pat_3_06_NA_0.txt" "BPatG_GebrM_LE
    _2006-10-16_5_W-pat_9_05_NA_0.txt" "BPatG_GebrM_LE_2007-08-20_5_W-pat_435_06_
    NA_0.txt" ...
##  $ url              : chr  "https://juris.bundesgerichtshof.de/cgi-bin/
    rechtsprechung/document.py?Gericht=bpatg&Art=en&Datum=2006&Sort=1&Se"| __
    truncated__ "https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/
    document.py?Gericht=bpatg&Art=en&Datum=2006&Sort=1&Se"| __truncated__ "https
    ://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=
    bpatg&Art=en&Datum=2006&Sort=1&Se"| __truncated__ "https://juris.
    bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bpatg&Art=en&
    Datum=2007&Sort=1&Se"| __truncated__ ...
##  $ gericht          : chr  "BPatG" "BPatG" "BPatG" "BPatG" ...
##  $ datum            : IDate, format: "2006-02-23" "2006-07-04" ...
##  $ entscheidungsjahr: int  2006 2006 2006 2007 2008 2009 2009 2009 2010 2010
    ...
##  $ entscheidung_typ : chr  "B" "B" "B" "B" ...
##  $ leitsatz         : chr  "LE" "LE" "LE" "LE" ...
##  $ senatsgruppe     : chr  "GebrM" "GebrM" "GebrM" "GebrM" ...
##  $ spruchkoerper_az : int  5 5 5 5 5 35 35 35 35 35 ...
##  $ registerzeichen  : chr  "W-pat" "W-pat" "W-pat" "W-pat" ...
##  $ verfahrensart    : chr  "Beschwerdeverfahren in Patentsachen,
    Gebrauchsmustersachen, Sortenschutzsachen, Markensachen" "Beschwerdeverfahren
     in Patentsachen, Gebrauchsmustersachen, Sortenschutzsachen, Markensachen" "
    Beschwerdeverfahren in Patentsachen, Gebrauchsmustersachen,
    Sortenschutzsachen, Markensachen" "Beschwerdeverfahren in Patentsachen,
    Gebrauchsmustersachen, Sortenschutzsachen, Markensachen" ...
##  $ eingangsnummer   : int  429 3 9 435 25 429 419 440 454 455 ...
##  $ eingangsjahr_az  : int  5 6 5 6 6 8 7 7 8 8 ...
##  $ eingangsjahr_iso : num  2005 2006 2005 2006 2006 ...
##  $ zusatz_az        : chr  NA NA NA NA ...
##  $ kollision        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ aktenzeichen     : chr  "5 W (pat) 429/05" "5 W (pat) 3/06" "5 W (pat)
    9/05" "5 W (pat) 435/06" ...
##  $ ecli             : chr  "ECLI:DE:BPatG:2006:230206B5Wpat429.05.0" "ECLI:DE:
    BPatG:2006:040706B5Wpat3.06.0" "ECLI:DE:BPatG:2006:161006B5Wpat9.05.0" "ECLI:
    DE:BPatG:2007:200807B5Wpat435.06.0" ...
##  $ bemerkung        : chr  "Leitsatzentscheidung" "Leitsatzentscheidung" "
    Leitsatzentscheidung" "Leitsatzentscheidung" ...
##  $ berichtigung     : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ wirkung          : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
##  $ ruecknahme       : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ erledigung       : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ zeichen          : int  14074 16369 25701 14446 22842 33066 12325 12550
    5438 5526 ...
##  $ tokens           : int  2202 2537 3810 2134 3244 4793 1693 1729 818 828 ...
##  $ typen            : int  642 735 1032 695 881 1087 598 605 321 331 ...
##  $ saetze           : int  140 131 213 90 137 194 76 78 61 62 ...
```

```
##  $ version          : chr  "2024-07-09" "2024-07-09" "2024-07-09" "2024-07-09"
     ...
##  $ doi_concept      : chr  "10.5281/zenodo.3954850" "10.5281/zenodo.3954850"
    "10.5281/zenodo.3954850" "10.5281/zenodo.3954850" ...
##  $ doi_version      : chr  "10.5281/zenodo.10849977" "10.5281/zenodo.10849977"
     "10.5281/zenodo.10849977" "10.5281/zenodo.10849977" ...
##  $ lizenz           : chr  "Creative Commons Zero 1.0 Universal" "Creative
    Commons Zero 1.0 Universal" "Creative Commons Zero 1.0 Universal" "Creative
    Commons Zero 1.0 Universal" ...
##  - attr(*, ".internal.selfref")=<externalptr>
##  - attr(*, "sorted")= chr "doc_id"
```

5.2 Hinweise

- **Abweichende Codierung der Registerzeichen** — Die Registerzeichen wurden gekürzt und Sonderzeichen entfernt um die Arbeit mit ihnen zu vereinfachen. Beachten Sie bitte die Gegenüberstellungstabelle unter Punkt 7.
- **Fehlende Werte** sind immer mit »NA« codiert.
- **Strings** können grundsätzlich alle in UTF-8 definierten Zeichen (insbesondere Buchstaben, Zahlen und Sonderzeichen) enthalten.

5.3 Erläuterung der Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV) Dateiname der extrahierten TXT-Datei.
text	String	(Nur CSV) Volltext der Entscheidung. Über Zeilengrenzen getrennte Wörter sind zusammengefügt.
url	String	(Nur CSV) Link zum Volltext der Entscheidung in der Datenbank des BPatG.
gericht	String	Name des Gerichts. Es ist nur der Wert »BPatG« vergeben. Dies ist der ECLI-Code für »Bundespatentgericht«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Datum (ISO)	Datum der Entscheidung im Format YYYY-MM-DD (ISO-8601).
entscheidungsjahr	Natürliche Zahl	(Nur CSV) Jahr der Entscheidung im Format YYYY (ISO-8601).
entscheidung_typ	String	Der Typ der Entscheidung. »B« für Beschluss, »U« für Urteil.
leitsatz	String	Ob es sich um eine Leitsatzentscheidung handelt. Der Wert ist entweder »LE« oder »NA«.

(continued)

Variable	Typ	Erläuterung
senatsgruppe	String	Senatsgruppe des Spruchkörpers. Jeder Senat is aufgrund seines speziellen Aufgabengebiets einer Senatsgruppe zugeteilt. Die verschiedenen Senatsgruppen und ihre Codierung sind im Abschnitt 6 erläutert.
spruchkoerper_az	String	Spruchkörper, wie er im jeweiligen amtlichen Aktenzeichen gelistet ist. Beim Bundespatentgericht handelt es sich um die Nummer des Senats. Aufgrund häufiger Senatsumbildungen liegen Entscheidungen vieler Senate vor, die aktuell nicht am Bundespatentgericht aktiv sind.
registerzeichen	String	Das amtliche Registerzeichen. Eine Erläuterung der Abkürzungen findet sich im Abschnitt 7.
verfahrensart	String	Die Verfahrensart, auf die das Registerzeichen hinweist. Siehe auch Abschnitt 7.
eingangsnummer	Natürliche Zahl	Eingangsnummer. Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine fortlaufene Nummer (Ordinalzahl) in der Reihenfolge ihres Eingangs, um sie voneinander abzugrenzen.
eingangsjahr_az	Natürliche Zahl	Eingangsjahr laut Aktenzeichen. Das Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Natürliche Zahl	Eingangsjahr im Format YYYY-MM-DD (ISO-8601).
zusatz_az	String	Optionaler Zusatz zum Aktenzeichen. Entscheidungen, die ein Europäisches Patent betreffen, enthalten den Zusatz »EP« (neuer) oder »EU« (älter). Alle anderen Entscheidungen sind mit »NA« markiert. Der Zusatz ist, wenn nicht »NA«, Teil des amtlichen Aktenzeichens.
kollision	Natürliche Zahl	Eine von mir vergebene Kollisionsnummer, falls die Dateinamen ansonsten identisch wären. Nicht vom BPatG vergeben!
aktenzeichen	String	Das amtliche Aktenzeichen im Format [senatsnummer] [registerzeichen] [eingangsnummer] / [eingangsjahr] [ggf. zusatz_az]

(continued)

Variable	Typ	Erläuterung
ecli	String	European Case Law Identifier (ECLI) der Entscheidung. Die ECLI wurde von mir berechnet und kann in Einzelfällen von der offiziell vergebenen ECLI abweichen. Das ist vor allem dann der Fall, wenn eine am gleichen Tag Entscheidungen mit dem gleichen Aktenzeichen ergangen sind und von mir eine Kollisionsnummer außer 0 vergeben wurde.
bemerkung	String	Bemerkung zur Entscheidung, wie sie in der Datenbank des BPatG dokumentiert ist.
berichtigung	Logisch	Ob das Dokument eine Berichtigung darstellt. Entweder »TRUE« oder »FALSE«. Aus der Variable »bemerkung« mit regular expressions extrahiert.
wirkung	Logisch	Ob die Entscheidung Wirkung entfaltet. Entweder »TRUE« oder »FALSE«. Aus der Variable »bemerkung« mit regular expressions extrahiert.
ruecknahme	Logisch	Ob die Klage zurückgenommen wurde. Entweder »TRUE« oder »FALSE«. Aus der Variable »bemerkung« mit regular expressions extrahiert.
erledigung	Logisch	Ob die Entscheidung erledigt ist. Entweder »TRUE« oder »FALSE«. Aus der Variable »bemerkung« mit regular expressions extrahiert.
zeichen	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Zeichen eines Dokumentes.
tokens	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.

(continued)

Variable	Typ	Erläuterung
typen	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl *einzigartiger* Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung und Typenzählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
saetze	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Sätze. Die Definition entspricht in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail allerdings sehr komplex und in »Unicode Standard: Annex No 29« beschrieben. Diese Zahl kann je nach Software und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Zählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
version	Datum (ISO)	(Nur CSV-Datei) Die Versionsnummer des Datensatzes im Format YYYY-MM-DD (Langform nach ISO-8601). Die Versionsnummer entspricht immer dem Datum an dem der Datensatz erstellt und die Daten von der Webseite des Gerichts abgerufen wurden.
doi_concept	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) des Gesamtkonzeptes des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer die **aktuellste Version** des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.

(continued)

Variable	Typ	Erläuterung
doi_version	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) der **konkreten Version** des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer diese konkrete Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.
lizenz	String	Die Lizenz für den Gesamtdatensatz. In diesem Datensatz immer »Creative Commons Zero 1.0 Universal«.

6 Senatsgruppen am Bundespatentgericht

Codierung	Senatsgruppe
GebrM	Gebrauchsmuster-Beschwerdesenat
JurBeschw	Juristischer Beschwerdesenat
JurBeschwNichtigkeit	Juristischer Beschwerdesenat und Nichtigkeitssenat
Marken	Marken-Beschwerdesenat
MarkenDesign	Marken- und Design-Beschwerdesenat
Nichtigkeit	Nichtigkeitssenat
Sortensch	Sortenschutz-Beschwerdesenat
TechnBeschw	Technischer Beschwerdesenat

7 Registerzeichen am Bundespatentgericht

Die im Datensatz enthaltenen Registerzeichen wurden jeweils um die runden Klammern bereinigt, um Probleme bei der Nutzung unter Windows zu vermeiden.

Register-zeichen	Codierung	Erläuterung
Li	Li	Zwangslizenzverfahren
LiQ	LiQ	Einstweilige Verfügungen in Zwangslizenzverfahren
Ni	Ni	Patentnichtigkeitsverfahren
W-(pat)	W-pat	Beschwerdeverfahren in Patentsachen, Gebrauchsmustersachen, Sortenschutzsachen, Markensachen
ZA-(Pat)	ZA-pat	Verfahren über Anträge außerhalb anhängiger Patentsachen

8 Linguistische Kennzahlen

8.1 Erläuterung der Kennzahlen und Diagramme

Zur besseren Einschätzung des inhaltlichen Umfangs des Korpus dokumentiere ich an dieser Stelle die Verteilung der Werte für einige klassische linguistische Kennzahlen:

Kennzahl	Definition
Zeichen	Zeichen entsprechen grob den *Graphemen*, den kleinsten funktionalen Einheiten in einem Schriftsystem. Beispiel: das Wort »Richterin« besteht aus 9 Zeichen.
Tokens	Eine beliebige Zeichenfolge, getrennt durch whitespace-Zeichen, d.h. ein Token entspricht in der Regel einem »Wort«, kann aber auch Zahlen, Sonderzeichen oder sinnlose Zeichenfolgen enthalten, weil es rein syntaktisch berechnet wird.
Typen	Einzigartige Tokens. Beispiel: wenn das Token »Verfassungsrecht« zehnmal in einer Entscheidung vorhanden ist, wird es als ein Typ gezählt.
Sätze	Entsprechen in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail aber sehr komplex und in »Unicode Standard: Annex No 29« beschrieben.

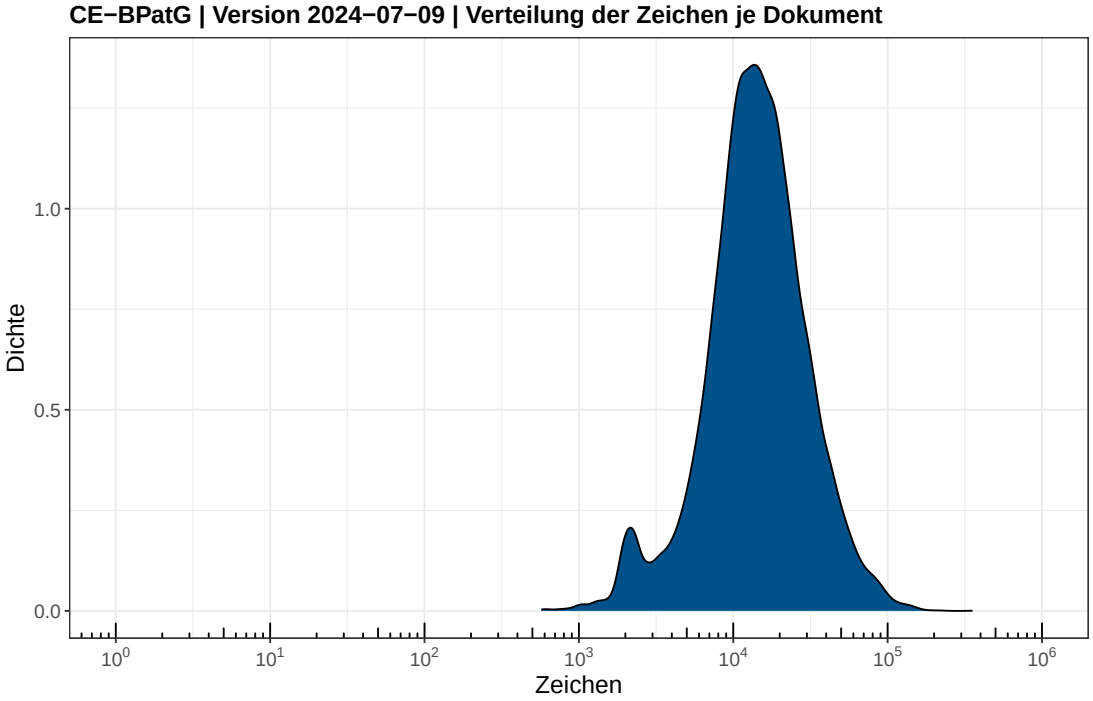
Es handelt sich bei den Diagrammen jeweils um »Density Charts«, die sich besonders dafür eignen die Schwerpunkte von Variablen mit stark schwankenden numerischen Werten zu visualisieren. Die Interpretation ist denkbar einfach: je höher die Kurve, desto dichter sind in diesem Bereich die Werte der Variable. Der Wert der y-Achse kann außer Acht gelassen werden, wichtig sind nur die relativen Flächenverhältnisse und die x-Achse.

Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

8.2 Werte der Kennzahlen

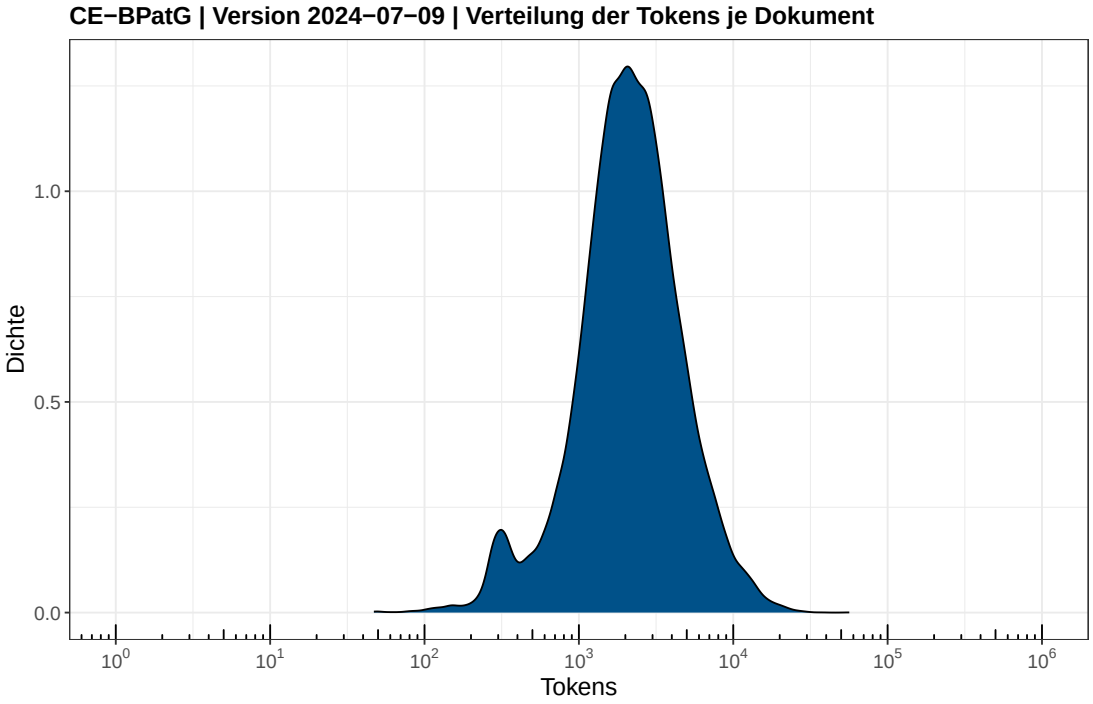
Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	565,970,711	571	9,106	14,154.5	18,336.38	22,378.75	354,918
tokens	86,127,153	47	1,334	2,128.0	2,790.36	3,419.75	56,435
typen	859,681	41	505	685.0	764.63	935.00	6,589
saetze	3,232,263	3	52	82.0	104.72	128.00	1,214

8.3 Verteilung Zeichen



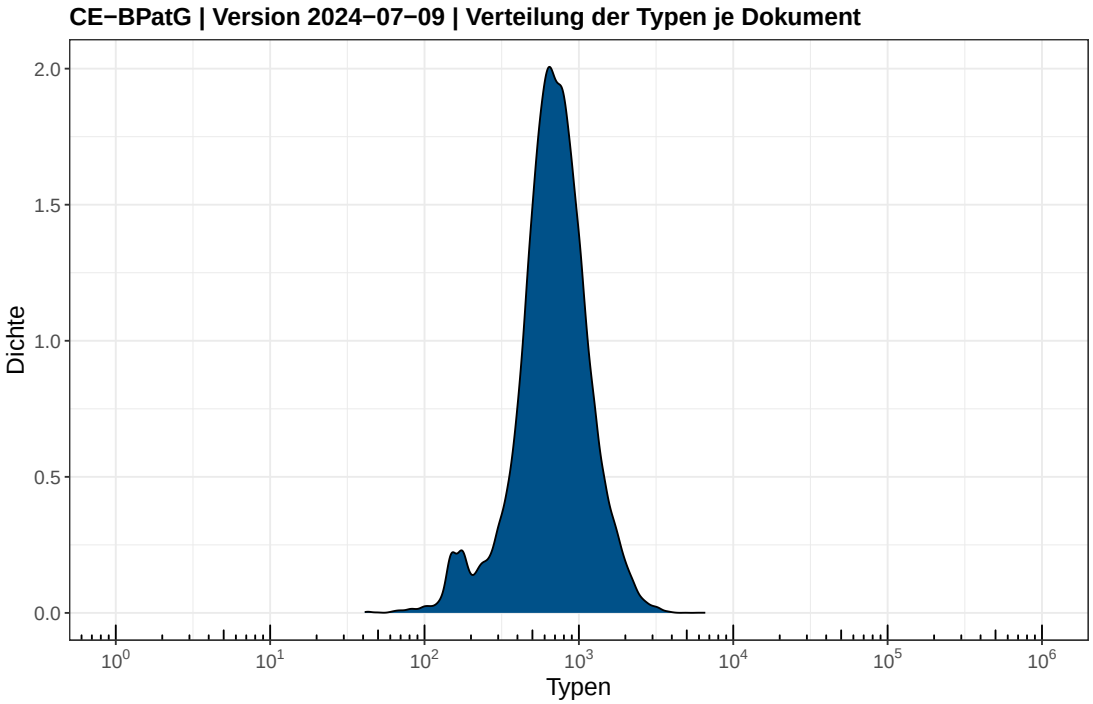
Fobbe | DOI: 10.5281/zenodo.10849977

8.4 Verteilung Tokens



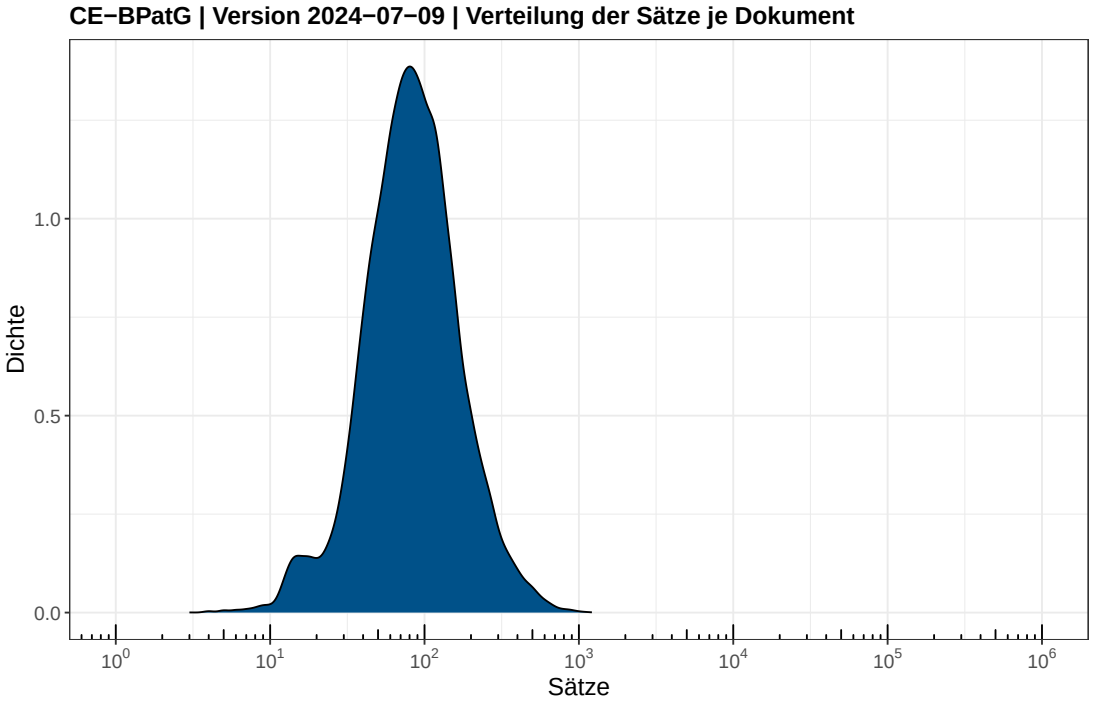
Fobbe | DOI: 10.5281/zenodo.10849977

8.5 Verteilung Typen



Fobbe | DOI: 10.5281/zenodo.10849977

8.6 Verteilung Sätze



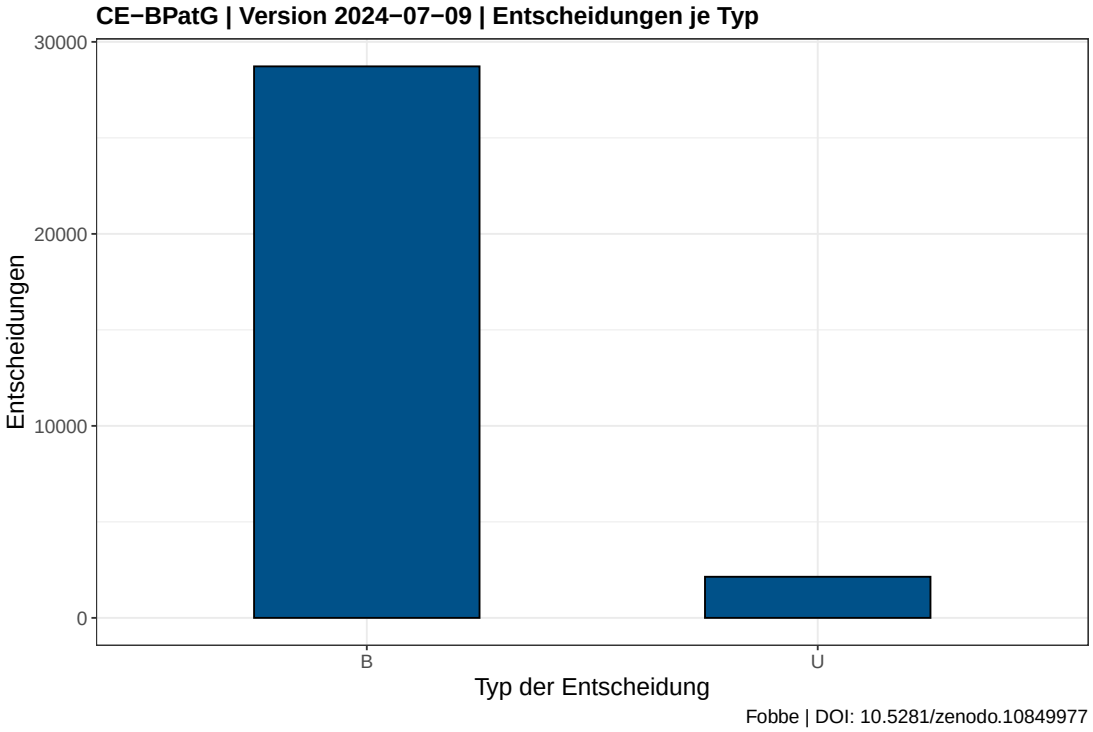
Fobbe | DOI: 10.5281/zenodo.10849977

9 Inhalt des Korpus

9.1 Zusammenfassung

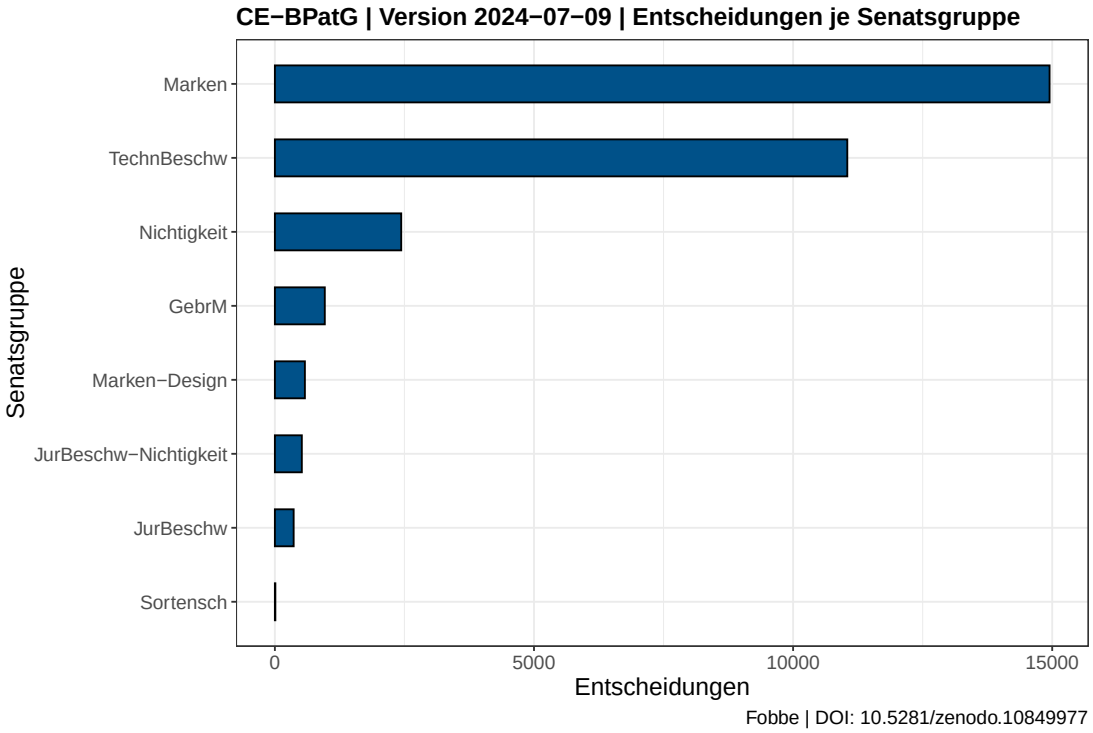
Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	25	2000	2004	2008	2008.92	2014	2024
eingangsjahr_iso	30	1977	2002	2005	2007.20	2012	2024
eingangsnummer	643	1	25	64	148.08	236	814

9.2 Nach Typ der Entscheidung



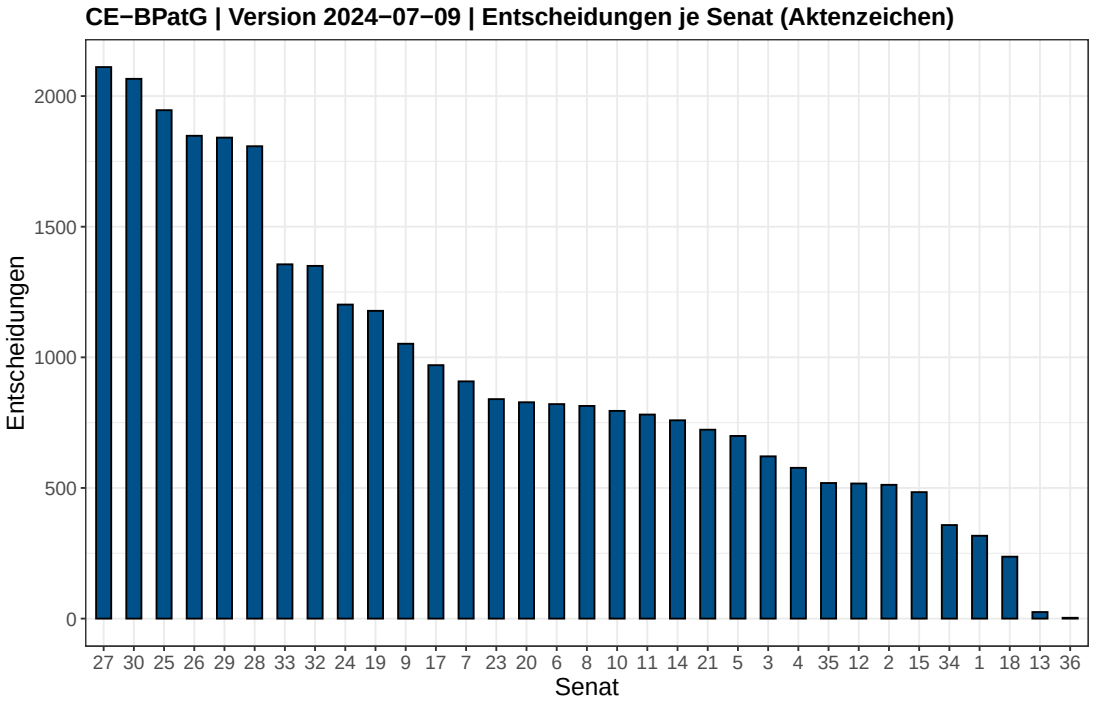
Typ	Entscheidungen	% Gesamt	% Kumulativ
B	28725	93.06	93.06
U	2141	6.94	100.00
Total	30866	100.00	100.00

9.3 Nach Senatsgruppe



Senatsgruppe	Entscheidungen	% Gesamt	% Kumulativ
GebrM	963	3.12	3.12
JurBeschw	362	1.17	4.29
JurBeschw-Nichtigkeit	520	1.68	5.98
Marken	14949	48.43	54.41
Marken-Design	579	1.88	56.29
Nichtigkeit	2438	7.90	64.18
Sortensch	9	0.03	64.21
TechnBeschw	11046	35.79	100.00
Total	30866	100.00	100.00

9.4 Nach Spruchkörper (Aktenzeichen)

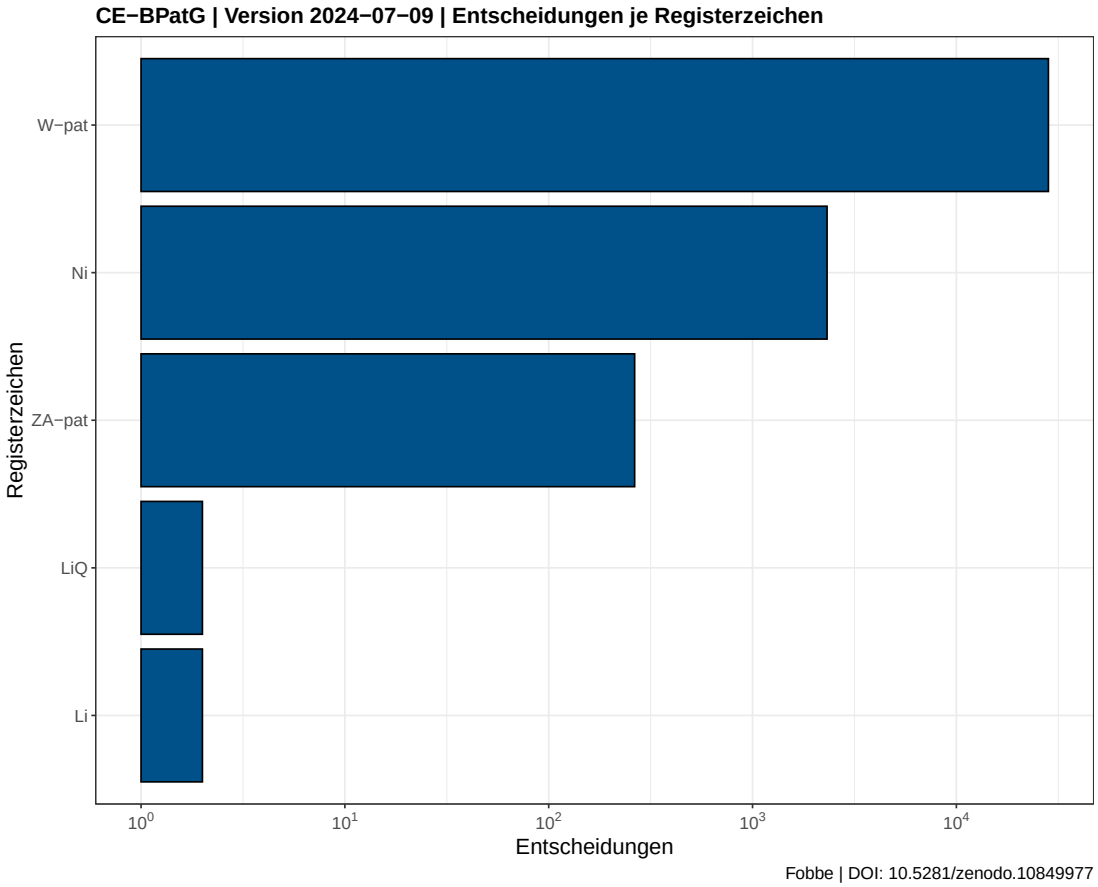


Senat	Entscheidungen	% Gesamt	% Kumulativ
1	317	1.03	1.03
2	512	1.66	2.69
3	621	2.01	4.70
4	577	1.87	6.57
5	699	2.26	8.83
6	821	2.66	11.49
7	908	2.94	14.43
8	814	2.64	17.07
9	1052	3.41	20.48
10	795	2.58	23.05
11	781	2.53	25.58
12	517	1.67	27.26
13	25	0.08	27.34
14	759	2.46	29.80
15	484	1.57	31.37
17	970	3.14	34.51
18	237	0.77	35.28

(continued)

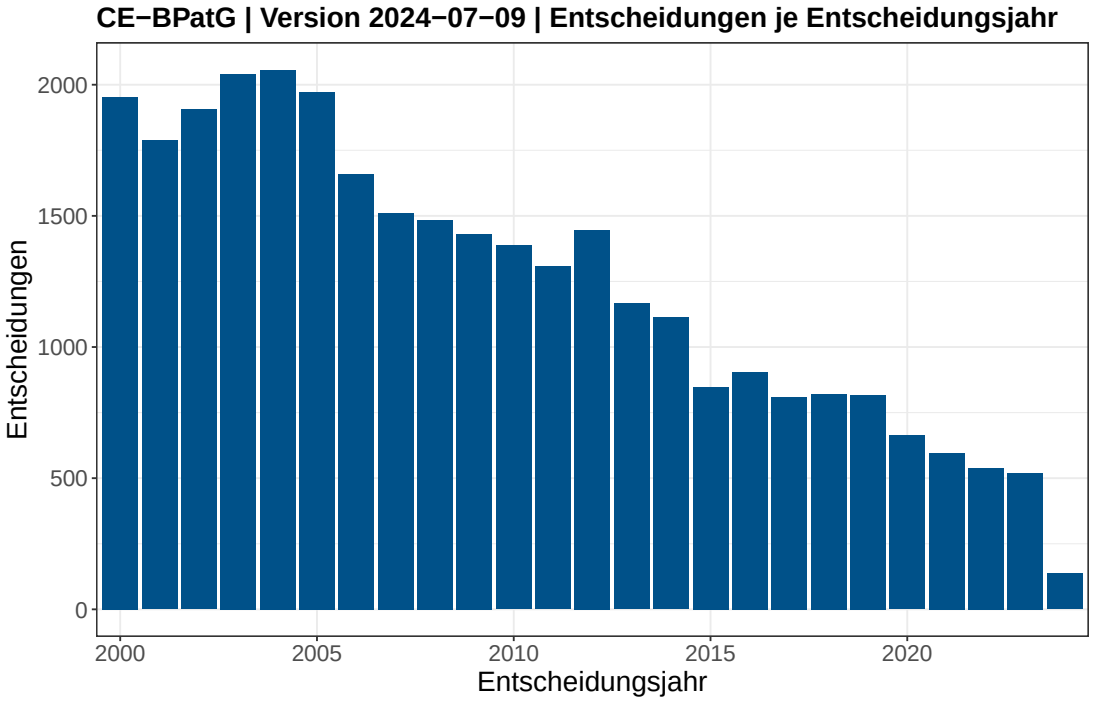
Senat	Entscheidungen	% Gesamt	% Kumulativ
19	1178	3.82	39.09
20	828	2.68	41.78
21	723	2.34	44.12
23	840	2.72	46.84
24	1202	3.89	50.74
25	1946	6.30	57.04
26	1848	5.99	63.03
27	2111	6.84	69.87
28	1808	5.86	75.72
29	1841	5.96	81.69
30	2066	6.69	88.38
32	1350	4.37	92.76
33	1356	4.39	97.15
34	358	1.16	98.31
35	519	1.68	99.99
36	3	0.01	100.00
Total	30866	100.00	100.00

9.5 Nach Registerzeichen



Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
Li	2	0.01	0.01
LiQ	2	0.01	0.01
Ni	2320	7.52	7.53
W-pat	28278	91.62	99.14
ZA-pat	264	0.86	100.00
Total	30866	100.00	100.00

9.6 Nach Entscheidungsjahr

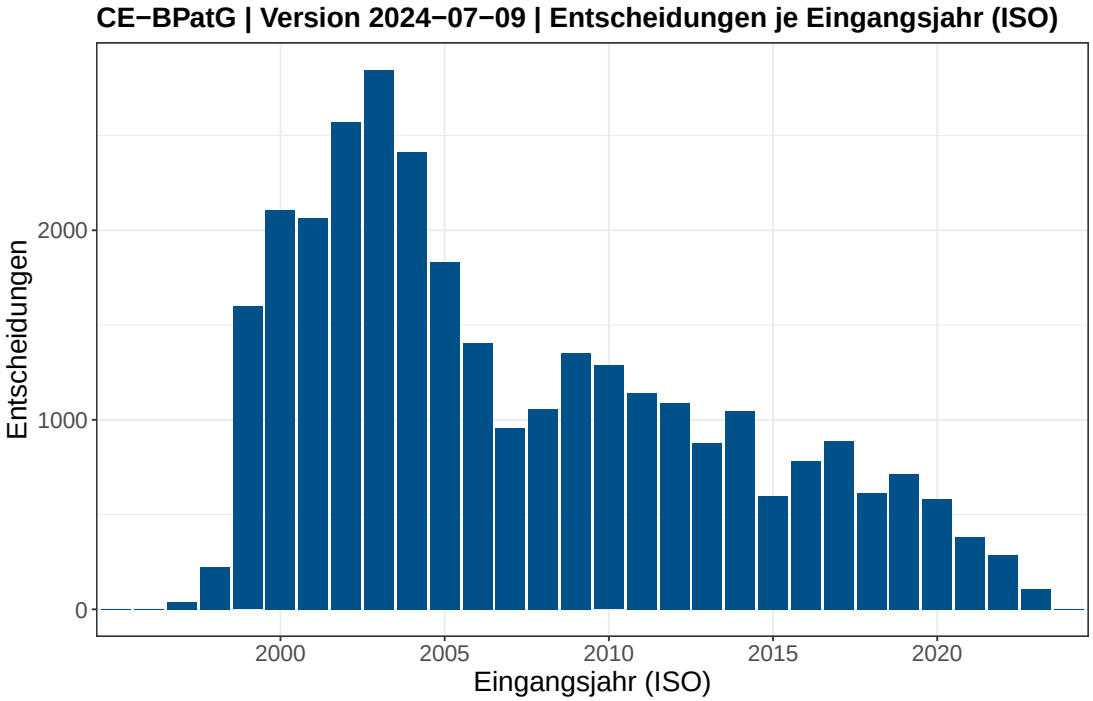


Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2000	1952	6.32	6.32
2001	1787	5.79	12.11
2002	1906	6.18	18.29
2003	2040	6.61	24.90
2004	2057	6.66	31.56
2005	1971	6.39	37.95
2006	1659	5.37	43.32
2007	1510	4.89	48.21
2008	1484	4.81	53.02
2009	1430	4.63	57.66
2010	1387	4.49	62.15
2011	1307	4.23	66.38
2012	1446	4.68	71.07
2013	1166	3.78	74.85
2014	1113	3.61	78.45
2015	847	2.74	81.20
2016	905	2.93	84.13

(continued)

Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2017	809	2.62	86.75
2018	820	2.66	89.41
2019	816	2.64	92.05
2020	663	2.15	94.20
2021	595	1.93	96.13
2022	539	1.75	97.87
2023	521	1.69	99.56
2024	136	0.44	100.00
Total	30866	100.00	100.00

9.7 Nach Eingangsjahr (ISO)

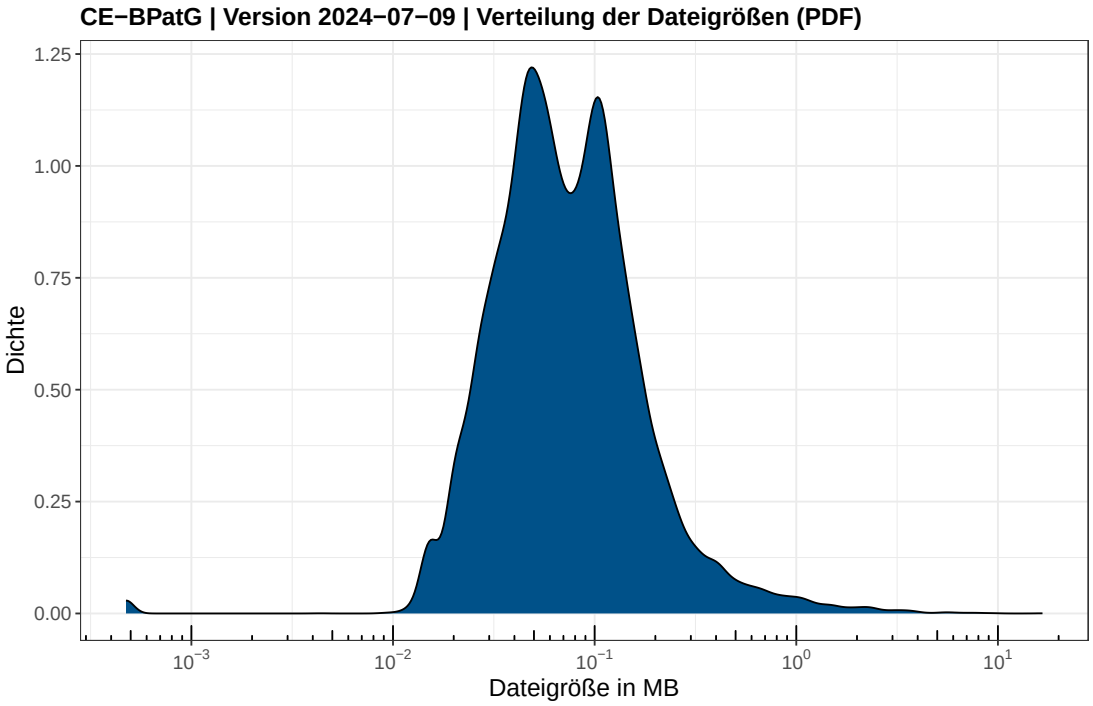


Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
1977	1	0.00	0.00
1996	2	0.01	0.01
1997	37	0.12	0.13
1998	225	0.73	0.86
1999	1598	5.18	6.04
2000	2108	6.83	12.87
2001	2064	6.69	19.55
2002	2571	8.33	27.88
2003	2847	9.22	37.11
2004	2414	7.82	44.93
2005	1833	5.94	50.87
2006	1407	4.56	55.42
2007	956	3.10	58.52
2008	1058	3.43	61.95
2009	1352	4.38	66.33
2010	1286	4.17	70.50
2011	1142	3.70	74.19

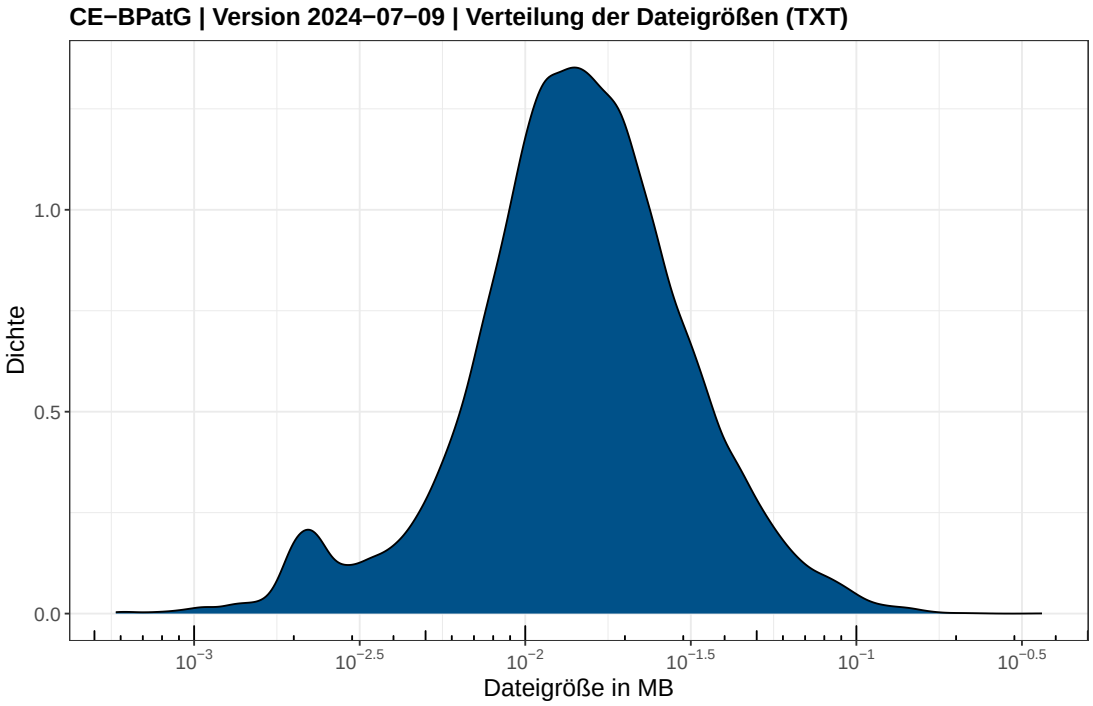
(continued)

Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
2012	1087	3.52	77.72
2013	878	2.84	80.56
2014	1045	3.39	83.95
2015	597	1.93	85.88
2016	781	2.53	88.41
2017	890	2.88	91.29
2018	614	1.99	93.28
2019	714	2.31	95.60
2020	581	1.88	97.48
2021	381	1.23	98.71
2022	287	0.93	99.64
2023	109	0.35	100.00
2024	1	0.00	100.00
Total	30866	100.00	100.00

10 Dateigrößen



Fobbe | DOI: 10.5281/zenodo.10849977



Fobbe | DOI: 10.5281/zenodo.10849977

11 Kryptographische Signaturen

11.1 Zwei-Phasen-Signatur

Die Integrität und Echtheit der einzelnen Archive des Datensatzes sind durch eine Zwei-Phasen-Signatur sichergestellt.

In **Phase I** werden während der Kompilierung für jedes ZIP-Archiv, das Codebook und die Robustness Checks Hash-Werte in zwei verschiedenen Verfahren (SHA2-256 und SHA3-512) berechnet und in einer CSV-Datei dokumentiert.

In **Phase II** werden diese CSV-Datei und der Compilation Report mit meinem persönlichen geheimen GPG-Schlüssel signiert. Dieses Verfahren stellt sicher, dass die Kompilierung von jedermann durchgeführt werden kann, insbesondere im Rahmen von Replikationen, die persönliche Gewähr für Ergebnisse aber dennoch vorhanden bleibt.

11.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen und der Compilation Report sind mit meiner persönlichen GPG-Signatur versehen. Der mit dieser Version korrespondierende Public Key ist sowohl mit dem Datensatz als auch mit dem Source Code hinterlegt. Er hat folgende Kenndaten:

Name: Sean Fobbe (fobbe-data@posteo.de)

Fingerabdruck: FE6F B888 F0E5 656C 1D25 3B9A 50C4 1384 F44A 4E42

12 Changelog

12.1 Version 2024-07-09

- LIZENZÄNDERUNG: Source Code jetzt unter GNU General Public License Version 3 (GPLv3) oder später lizenziert
- Vollständige Aktualisierung der Daten
- R-Version auf 4.4.0 aktualisiert (wegen CVE-2024-27322)
- Vereinfachung der Repository-Struktur mit Ordner etc/ für Config Files
- Anpassung von Docker Compose File an Debian 11
- Docker Zeitzone auf Berlin eingestellt
- Aktualisierung von Public GPG Key im Repository
- Bei Auswahl maximaler Cores werden max(cores)-1 benutzt
- Neues Profiling der Größe von PDF- und TXT-Dateien
- Python Toolchain entfernt
- Konvertierung von PDF zu TXT wird bei Fehlern nicht mehr unterbrochen (u.a. um fälschlich ausgelieferte HTML-Dateien zu übergehen)
- Variable "entscheidung_typ" im Codebook dokumentiert
- Pipeline prüft nun automatisch ob alle Variablen im Datensatz auch im Codebook dokumentiert wurden

12.2 Version 2023-04-02

- Vollständige Aktualisierung der Daten
- Gesamte Laufzeitumgebung mit Docker versionskontrolliert
- Aktenzeichen aus dem Eingangszeitraum 2000 bis 2009 nun korrekt mit führender Null formatiert (z.B. 1 BvR 44/02 statt 1 BvR 44/2)
- Vereinfachung der Konfigurationsdatei
- Run- und Delete-Skripte aktualisiert
- Neue Funktion für automatischen clean run (Löschung aller Zwischenergebnisse)
- Neuorganisation des Repositories
- Inhalt des ZIP-Archivs mit dem Source Code orientiert sich nun an der Versionskontrolle mit Git und enthält auch die gesamte Git-Historie
- Proto-Package Mono-Repo entfernt, alle Funktionen nun fest projektbasiert versionskontrolliert
- Update der Download-Funktion
- Überflüssige Warnung in f.future_lingsummarize-Funktion entfernt
- Zusätzliche Unit-Tests
- Alle Roh-Dateien werden nun im Ordner "files/" gespeichert
- Verbesserung des Robustness Check Reports
- Verbesserung des Codebooks
- Alle Diagramme neu nummeriert
- Verbesserte Formatierung von Profiling, Warnungen und Fehlermeldungen im Compilation Report
- README im Hinblick auf Docker überarbeitet
- Alle Zwischenergebnisse der Pipeline werden automatisch im Ordner "output/" archiviert
- Umfang der Datenbankabfrage ist nun vollständig automatisiert
- Zwischenergebnisse werden im qs-Format gespeichert um Speicherplatz zu sparen

12.3 Version 2022-07-11

- Vollständige Aktualisierung der Daten
- Neuer Entwurf des gesamten Source Code im {targets} Framework
- Veröffentlichung des Source Codes

12.4 Version 2020-07-20

- Erstveröffentlichung

13 Parameter für strenge Replikationen

```
## [1] "OpenSSL 3.0.2 15 Mar 2022 (Library: OpenSSL 3.0.2 15 Mar 2022)"
```

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.4 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.20.so;
    LAPACK version 3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Europe/Berlin
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] future.apply_1.11.2 future_1.33.2       quanteda_4.0.2
##  [4] readtext_0.91       data.table_1.15.4   scales_1.3.0
##  [7] ggraph_2.2.1        ggplot2_3.5.1       pdftools_3.4.0
## [10] kableExtra_1.4.0    knitr_1.46          testthat_3.2.1.1
## [13] rvest_1.0.4         httr_1.4.7          mgsub_1.7.3
## [16] zip_2.3.1           fs_1.6.4            RcppTOML_0.2.2
## [19] tarchetypes_0.9.0   targets_1.7.0
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.1    viridisLite_0.4.2   dplyr_1.1.4
##  [4] farver_2.1.1        viridis_0.6.5       fastmap_1.1.1
##  [7] tweenr_2.0.3        stringfish_0.16.0   digest_0.6.35
## [10] base64url_1.4       lifecycle_1.0.4     secretbase_0.5.0
## [13] qpdf_1.3.3          waldo_0.5.2         processx_3.8.4
## [16] magrittr_2.0.3      compiler_4.4.0      rlang_1.1.3
## [19] tools_4.4.0         igraph_2.0.3        utf8_1.2.4
## [22] yaml_2.3.8          labeling_0.4.3      askpass_1.2.0
## [25] stopwords_2.3       graphlayouts_1.1.1  xml2_1.3.6
## [28] pkgload_1.3.4       withr_3.0.0         purrr_1.0.2
## [31] desc_1.4.3          grid_4.4.0          polyclip_1.10-6
## [34] fansi_1.0.6         colorspace_2.1-0    globals_0.16.3
## [37] MASS_7.3-60.2       tinytex_0.51        cli_3.6.2
## [40] rmarkdown_2.26      generics_0.1.3      RcppParallel_5.1.7
## [43] rstudioapi_0.16.0   RApiSerialize_0.1.2 cachem_1.0.8
## [46] ggforce_0.4.2       stringr_1.5.1       parallel_4.4.0
```

```
## [49] vctrs_0.6.5          Matrix_1.7-0         callr_3.7.6
## [52] ggrepel_0.9.5        listenv_0.9.1        systemfonts_1.0.6
## [55] tidyr_1.3.1          glue_1.7.0           parallelly_1.37.1
## [58] codetools_0.2-20     ps_1.7.6             stringi_1.8.4
## [61] gtable_0.3.5         munsell_0.5.1        tibble_3.2.1
## [64] pillar_1.9.0         htmltools_0.5.8.1    brio_1.1.5
## [67] R6_2.5.1             rprojroot_2.0.4      tidygraph_1.3.1
## [70] evaluate_0.23        lattice_0.22-6       qs_0.26.1
## [73] backports_1.4.1      memoise_2.0.1        Rcpp_1.0.12
## [76] fastmatch_1.1-4      svglite_2.1.3        gridExtra_2.3
## [79] xfun_0.43            pkgconfig_2.0.3
```

Literaturverzeichnis

Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2024. *Rmarkdown: Dynamic Documents for R*. https://github.com/rstudio/rmarkdown.

Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, and Toby Hocking. 2024. *Data.table: Extension of 'Data.frame'*. https://r-datatable.com.

Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in R Using Futures." *The R Journal* 13 (2): 208–27. https://doi.org/10.32614/RJ-2021-048.

———. 2024a. *Future.apply: Apply Function to Elements in Parallel Using Futures*. https://future.apply.futureverse.org.

———. 2024b. *Future: Unified Parallel and Distributed Processing in R for Everyone*. https://future.futureverse.org.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. https://doi.org/10.21105/joss.00774.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe. 2024. *Quanteda: Quantitative Analysis of Textual Data*. https://quanteda.io.

Csárdi, Gábor. 2024. *Zip: Cross-Platform Zip Compression*. https://github.com/r-lib/zip.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems: 1695. https://igraph.org.

Csárdi, Gábor, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller. 2024. *Igraph: Network Analysis and Visualization*. https://r.igraph.org/.

Eddelbuettel, Dirk. 2023. *RcppTOML: Rcpp Bindings to Parser for "Tom's Obvious Markup Language"*. http://dirk.eddelbuettel.com/code/rcpp.toml.html.

Ewing, Mark. 2021. *Mgsub: Safe, Multiple, Simultaneous String Substitution*.

Gagolewski, Marek. 2022. "stringi: Fast and Portable Character String Processing in R." *Journal of Statistical Software* 103 (2): 1–59. https://doi.org/10.18637/jss.v103.i02.

Gagolewski, Marek, Bartek Tartanus, others; Unicode, Inc., and others. 2024. *Stringi: Fast and Portable Character String Processing Facilities*. https://stringi.gagolewski.com/.

Landau, William Michael. 2021a. *Tarchetypes: Archetypes for Targets*.

———. 2021b. "The Targets R Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 6 (57): 2959. https://doi.org/10.21105/joss.02959.

———. 2024a. *Tarchetypes: Archetypes for Targets*. https://docs.ropensci.org/tarchetypes/.

———. 2024b. *Targets: Dynamic Function-Oriented Make-Like Declarative Pipelines*. https://docs.ropensci.org/targets/.

Ooms, Jeroen. 2023. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. https://docs.ropensci.org/pdftools/.

Pedersen, Thomas Lin. 2024. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. https://ggraph.data-imaginist.com.

Ushey, Kevin, and Hadley Wickham. 2024. *Renv: Project Environments*. https://rstudio.github.io/renv/.

Wickham, Hadley. 2024. *Rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.org/knitr/.

———. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://yihui.org/knitr/.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.

Zhu, Hao. 2024. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.