

CORPUS DER ENTSCHEIDUNGEN DES
BUNDESPATENTGERICHTS
(CE-BPatG)

CODEBOOK

Version 2020-07-20

DOI: [10.5281/zenodo.3954851](https://doi.org/10.5281/zenodo.3954851)

Titel	Corpus der Entscheidungen des Bundespatentgerichts
Abkürzung	CE-BPatG
Autor	Sean Fobbe
Version	2020-07-20
Download	https://doi.org/10.5281/zenodo.3954851
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Zitieren Sie diesen Datensatz bitte, wenn Sie ihn in einem akademischen Kontext benutzen. Beispielsweise mit diesem Zitiervorschlag:

Fobbe, Sean, Corpus der Entscheidungen des Bundespatentgerichts (CE-BPatG), Version 2020-07-20, Zenodo, DOI: <https://doi.org/10.5281/zenodo.3954851>

Digital Object Identifiers: Concept DOI und Version DOI

Die in diesem Codebook verwendete DOI ist eine »concept DOI« und bezieht sich auf das Gesamtkonzept des Datensatzes. Sie verlinkt daher immer automatisch die aktuellste Version. Für jede spezifische Version steht eine weitere »version DOI« zur Verfügung, die auf der Zenodo-Seite der jeweiligen Version abgerufen werden kann. Diese Seiten sind auch über Links von der aktuellsten Version erreichbar.

Urheberrecht

Der Datensatz und dieses Dokument sind unter einer **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication** Lizenz veröffentlicht. Ich stelle den Datensatz und das Codebook vollständig gemeinfrei und verzichte weltweit auf alle damit verbundenen Urheberrechte, einschließlich aller ähnlichen Rechte, soweit dies gesetzlich möglich ist.

Sie können die Werke kopieren, modifizieren, verteilen und aufführen ohne um Erlaubnis bitten zu müssen, selbst für kommerzielle Zwecke. Patente und Markenrechte bleiben von CC0 unberührt. CC0 hat auch keine Auswirkungen auf etwaige Datenschutz- oder Persönlichkeitsrechte. Jegliche Haftung für die Benutzung dieses Werkes ist ausgeschlossen, bis zu dem maximalen Umfang in dem dies gesetzlich möglich ist.

Wenn Sie diese Werke nutzen oder zitieren sollten Sie nicht den Eindruck erwecken, der Autor unterstütze ihre Nutzung.

Dies ist nur eine unverbindliche deutsche Zusammenfassung der Lizenz, den vollständigen und rechtsverbindlichen Lizenztext finden Sie hier:

<https://creativecommons.org/publicdomain/zero/1.0/legalcode>

Inhaltsverzeichnis

Inhaltsverzeichnis	3
1 Einführung	4
2 Datenquelle	4
3 Hinweis zum Urheberrecht an den Rohdaten	5
4 Metadaten	5
4.1 Schema	5
4.2 Beispiel	5
4.3 Nutzung mit R	6
4.4 Details zu den Variablen	6
4.5 Hinweis: Abweichende Codierung der Registerzeichen	7
5 Senatsgruppen am Bundespatentgericht	8
6 Registerzeichen am Bundespatentgericht	8
7 Inhalt	9
7.1 Nach Spruchkörper	9
7.2 Nach Senatsgruppe	10
7.3 Nach Registerzeichen	10
7.4 Nach Jahr	11
8 Disclaimer	11

1 Einführung

Die quantitative Analyse von juristischen Texten ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld. 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*¹ und *openJur*² verbieten ausdrücklich das maschinelle Auslesen der Rohdaten.³ 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 das, dass sie auch quantitativen Analysen zugänglich sein muss. Der Erstellung und Aufbereitung des Datensatzes liegen daher die Gedanken der Transparenz und wissenschaftlichen Reproduzierbarkeit zugrunde.

Zusätzlich zu den einfach maschinenlesbaren Formaten (TXT und CSV) sind die PDF-Rohdaten enthalten, damit Analysten gegebenenfalls ihre eigene Konvertierung vornehmen können. Die PDF-Rohdaten wurden inhaltlich nicht verändert und die Dateinamen nur geringfügig angepasst um die Lesbarkeit für Mensch und Maschine zu verbessern. Die TXT-Dateien wurden für alle Urteile mit **pdftotext 0.73.0** erstellt. Die Text-Dateien wurden anschließend mit **readtext 0.7.6** in **R 3.6.3** eingelesen und das CSV file mit dem Standard-Befehl **write.csv()** erstellt.

2 Datenquelle

Die Datenquelle ist: <https://www.bundespatentgericht.de/>

Der Datensatz enthält alle Entscheidungen die zum jeweiligen Stichtag (=Versionsnummer) auf der offiziellen Webseite des Bundespatentgerichts in dessen amtlicher Entscheidungsdatenbank abrufbar waren. Die Daten wurden unter Beachtung des Robot Exclusion Standard (RES) gesammelt.

Nutzer sollten zwei wichtige Einschränkungen beachten:

1. Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht (*publication bias*)
2. Entscheidungen sind erst ab dem Jahr 2000 enthalten

¹ <https://opinioiuris.de/>

² <https://openjur.de/>

³ Openjur beabsichtigt eine API anzubieten, diese war aber im Juli 2020 immernoch nicht verfügbar. Openjur ist seit 2008 in Betrieb.

3 Hinweis zum Urheberrecht an den Rohdaten

An den Entscheidungstexten und amtlichen Leitsätzen besteht gem. § 5 I UrhG kein Urheberrecht, da sie amtliche Werke sind. § 5 UrhG ist auf amtliche Datenbanken analog anzuwenden (BGH, Beschluss vom 28.09.2006 - I ZR 261/03, »Sächsischer Ausschreibungsdienst«). Alle eigenen Beiträge (z.B. durch Zusammenstellung und Anpassung der Metadaten) stelle ich gemäß einer *CC0 1.0 Universal Public Domain Lizenz* vollständig urheberrechtsfrei.

4 Metadaten

Die Metadaten in den Dateinamen sind größtenteils unverändert von den jeweiligen Datenbankeinträgen aus der amtlichen Datenbank des Bundespatentgerichts entnommen. Hinzugefügt wurden von mir der Gerichtsname, sowie Unter- und Trennstriche um die Maschinenlesbarkeit zu erleichtern.

Die Dateinamen enthalten Gerichtsname, Datum (nach ISO-8601, d.h. YYYY-MM-DD), die Senatsgruppe und das offizielle Aktenzeichen.

Die Typen der Variablen wurden mit *regular expressions* strikt validiert.

4.1 Schema

[Gericht]_[Datum]_[Senatsgruppe]_[SpruchkoerperAZ]_[Registerzeichen]_
[Ordinalzahl]_[Eingangsjahr]_[Suffix]

4.2 Beispiel

BPatG_2020-06-18_Nichtigkeit_4_Ni_6_19_EP

4.3 Nutzung mit R

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

```
txt.bpatg <- readtext("./CE-BPatG/*.txt",
                     docvarsfrom = "filenames",
                     docvarnames = c("Gericht",
                                     "Datum",
                                     "Senatsgruppe",
                                     "SpruchkoerperAZ",
                                     "Registerzeichen",
                                     "Ordinalzahl",
                                     "Eingangsjahr",
                                     "Suffix"),
                     dvsep = "_",
                     encoding = "UTF-8")
```

Die **CSV**-Datei ist mit **R** folgendermaßen einlesbar:

```
csv.bpatg <- read.csv("./Dateiname.csv")
```

Selbstverständlich lassen sich die Daten auch mit anderen Programmiersprachen (z.B. Python) und grafischen Programmen nutzen.

4.4 Details zu den Variablen

Variable	Typ	Erläuterung
Gericht	Alphabetisch	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	Das Datum der Entscheidung im Format YYYY-MM-DD (ISO-8601).
Senatsgruppe	Alphabetisch	Jeder Senat ist aufgrund seines speziellen Aufgabengebiets einer Senatsgruppe zugeteilt. Die verschiedenen Senatsgruppen und ihre Codierung sind unter Punkt 5 erläutert.

SpruchkoerperAZ	Alphanumerisch	Der 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	Alphabetisch	Das amtliche Registerzeichen. Eine Erläuterung der Abkürzungen findet sich unter Punkt 6.
Ordinalzahl	Natürliche Zahl	Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine fortlaufene Nummer (Ordinalzahl) in der Reihenfolge ihres Eingangs, um sie voneinander abzugrenzen.
Eingangsjahr	Natürliche Zahl	Das Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
Suffix	Alphabetisch	Entscheidungen, die ein Europäisches Patent betreffen, enthalten den Suffix »EP« (neuer) oder »EU« (älter). Alle anderen Entscheidungen sind mit »NA« markiert. Das Suffix ist, wenn nicht »NA«, Teil des amtlichen Aktenzeichens.

4.5 Hinweis: Abweichende Codierung der Registerzeichen

Windows verbietet gewisse Sonderzeichen in Dateinamen (z.B. runde Klammern). Die Registerzeichen wurden daher gekürzt und Sonderzeichen entfernt um Probleme für Windows-Nutzer zu vermeiden. Beachten Sie bitte die Gegenüberstellungstabelle unter dem Punkt 6.

5 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

6 Registerzeichen am Bundespatentgericht

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

Registerzeichen	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

7 Inhalt

7.1 Nach Spruchkörper

Senat	Entscheidungen	% Gesamt	% Kumulativ
1	241	0.84	0.84
2	465	1.62	2.46
3	566	1.97	4.44
4	508	1.77	6.21
5	658	2.29	8.5
6	735	2.56	11.07
7	854	2.98	14.04
8	739	2.58	16.62
9	975	3.4	20.02
10	795	2.77	22.79
11	720	2.51	25.3
12	402	1.4	26.71
13	25	0.09	26.79
14	698	2.43	29.23
15	483	1.68	30.91
17	896	3.12	34.04
18	177	0.62	34.65
19	1099	3.83	38.49
20	780	2.72	41.21
21	724	2.52	43.73
23	805	2.81	46.54
24	1202	4.19	50.73
25	1752	6.11	56.84
26	1691	5.9	62.74
27	2109	7.35	70.09
28	1644	5.73	75.82
29	1651	5.76	81.58
30	1801	6.28	87.86
32	1352	4.71	92.58
33	1356	4.73	97.3
34	358	1.25	98.55
35	413	1.44	99.99
36	2	0.01	100

<NA>	0	0	100
Total	28676	100	100

7.2 Nach Senatsgruppe

Senatsgruppe	Entscheidungen	% Gesamt	% Kumulativ
GebrM	856	2.99	2.99
JurBeschw	361	1.26	4.24
JurBeschwNichtigkeit	468	1.63	5.88
Marken	14248	49.69	55.56
MarkenDesign	310	1.08	56.64
Nichtigkeit	2064	7.2	63.84
Sortensch	8	0.03	63.87
TechnBeschw	10361	36.13	100
<NA>	0	0	100
Total	28676	100	100

7.3 Nach Registerzeichen

Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
Li	2	0.01	0.01
LiQ	2	0.01	0.01
Ni	1946	6.79	6.8
W-pat	26463	92.28	99.08
ZA-pat	263	0.92	100
<NA>	0	0	100
Total	28676	100	100

7.4 Nach Jahr

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2000	1951	6.8	6.8
2001	1790	6.24	13.05
2002	1910	6.66	19.71
2003	2046	7.13	26.84
2004	2060	7.18	34.02
2005	1970	6.87	40.89
2006	1663	5.8	46.69
2007	1511	5.27	51.96
2008	1483	5.17	57.13
2009	1425	4.97	62.1
2010	1395	4.86	66.97
2011	1309	4.56	71.53
2012	1450	5.06	76.59
2013	1167	4.07	80.66
2014	1119	3.9	84.56
2015	845	2.95	87.51
2016	905	3.16	90.66
2017	805	2.81	93.47
2018	824	2.87	96.35
2019	798	2.78	99.13
2020	250	0.87	100
<NA>	0	0	100
Total	28676	100	100

8 Disclaimer

Dieser Datensatz ist eine private wissenschaftliche Initiative und steht in keiner Verbindung zum Bundespatentgericht.