Extraktion von **ju**ristischen **Ver**weisstrukturen: Intertextualität in Gesetzen, Kommentaren und Gerichtsurteilen

Case law corpus 1.1.0

Das Korpus enthält anonymisierte deutsche Gerichtsentscheidungen aus deutschen Strafverfahren aus den Jahren 2015 bis Anfang 2020. Die Gerichtsentscheidungen werden vom Bundesgerichtshof (BGH) veröffentlicht. Die Gerichtsentscheidungen wurden direkt von der Website der Rechtsprechungsdatenbank gecrawlt und stehen im HTML-Format für die weitere Arbeit zur Verfügung. Die Quelle für die Dokumente im Rechtsprechungskorpus finden Sie unter: hrr-strafrecht.de.

Format der Dateien: Da die strafrechtlichen Gerichtsentscheidungen von einer Webseite gecrawlt wurden, liegen die Originaldateien im HTML-Format (PHP) vor.

Struktur der Dateien: Die Urteile wurden als vollständige Seiten gesichert, d.h. die Menüstruktur, die Metadaten und die HTML-Struktur stehen zur Verfügung. Dubletten in den Urteilen wurden entfernt.

Lizenz: Siehe hrr-strafrecht.de

Umfang: 4.250 Dokumente

Größe: Ca. 118 MB

Datum: Juli 2021

Projektleiter: Prof. Dr. Christian Wartena

Verantwortlich Rechtswissenschaften: Prof. Dr. Fabian Schmieder

Projektlaufzeit: 01.10.2019 bis 30.09.2022

Projektförderung: Niedersächsischen Ministeriums für Wissenschaft und Kultur, Förderinitiative: Niedersächsisches Vorab, Förderlinie: Geistes- und Kulturwissenschaften – digital: Forschungschancen, Methodenentwicklung und Reflexionspotenziale

Projektwebseite: http://textmining.wp.hs-hannover.de/juver.html

Kontakt:
Hochschule Hannover
Prof. Dr. Christian Wartena
Expo Plaza 12
30539 Hannover
E-Mail: christian.wartena@hs-hannover.de