



Foto: UHH/Lutsch

Magnus Bender, Colloquium, 02.10.2024

Toward better Subjective Content Descriptions (SCDs): A Spiral of Estimation, Enrichment, Usage, and Improvement

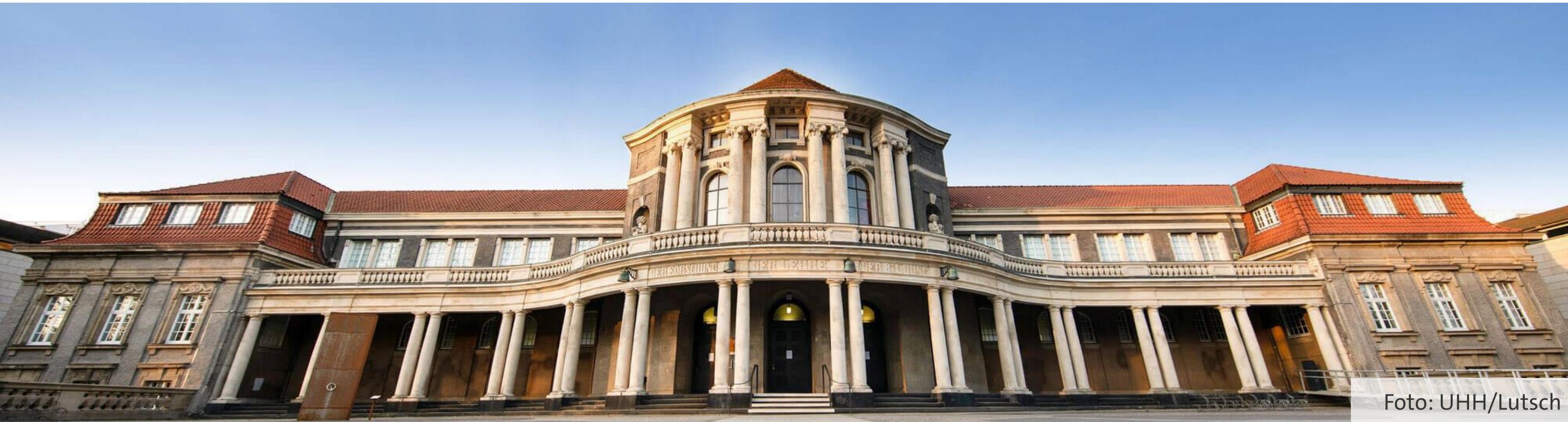
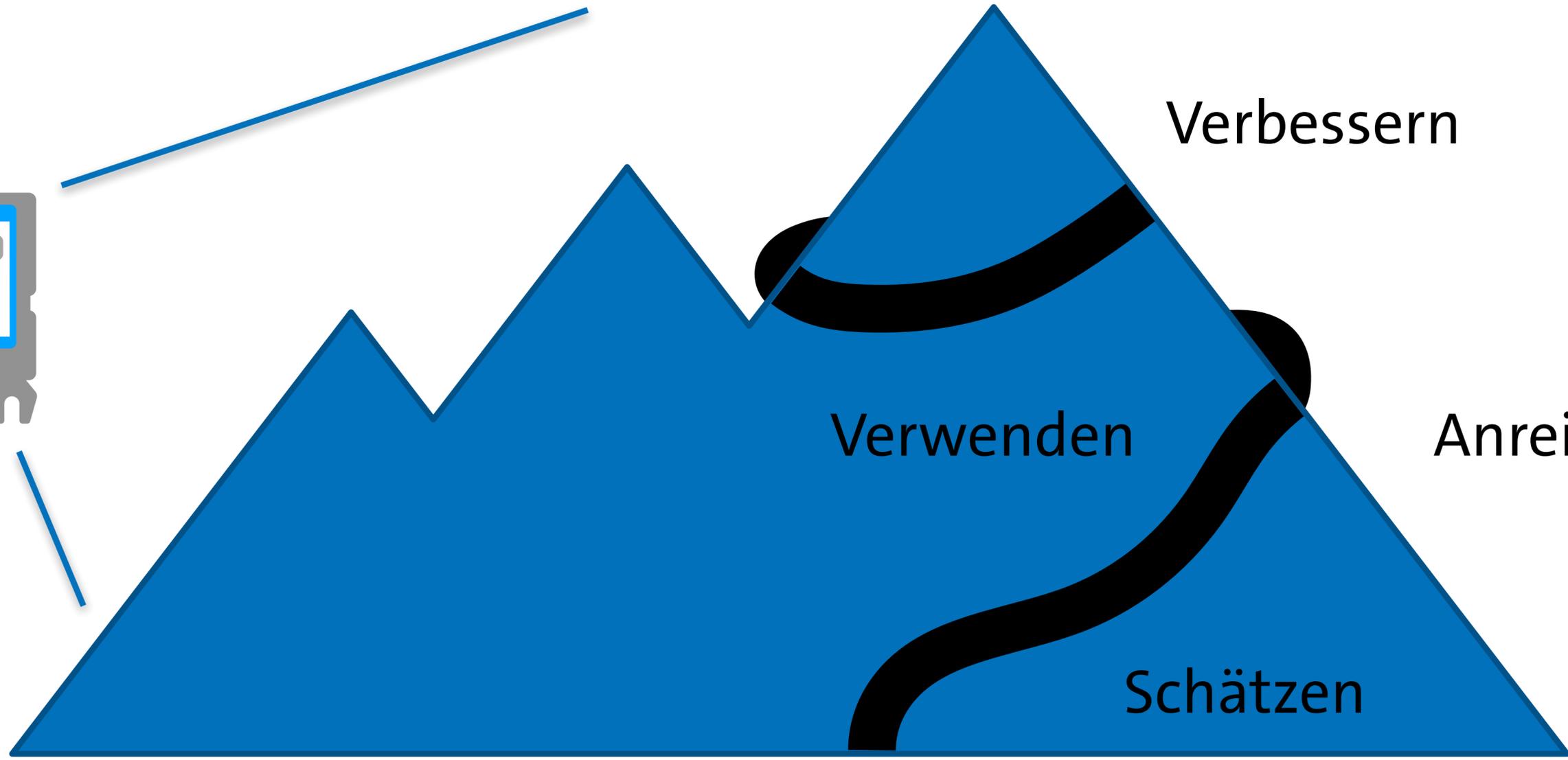
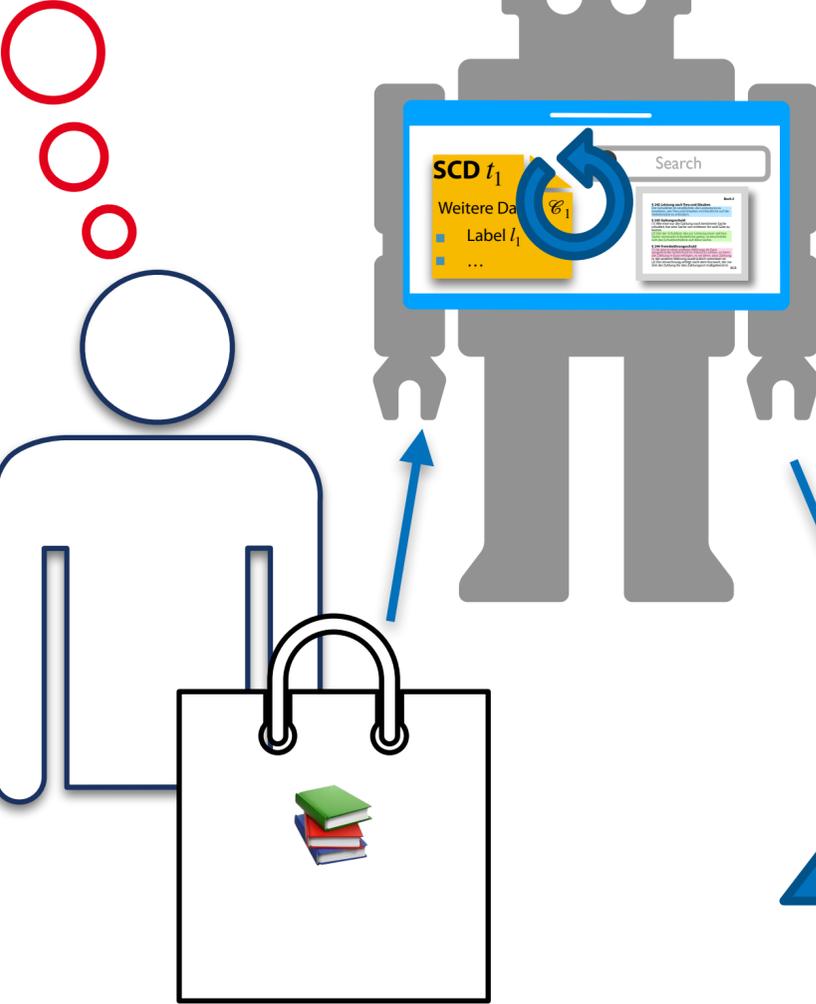


Foto: UHH/Lutsch

Magnus Bender, Kolloquium, 02.10.2024

Unterwegs zu besseren SCDs: Eine Spirale aus schätzen, anreichern, verwenden und verbessern

Persönlicher
Blickwinkel



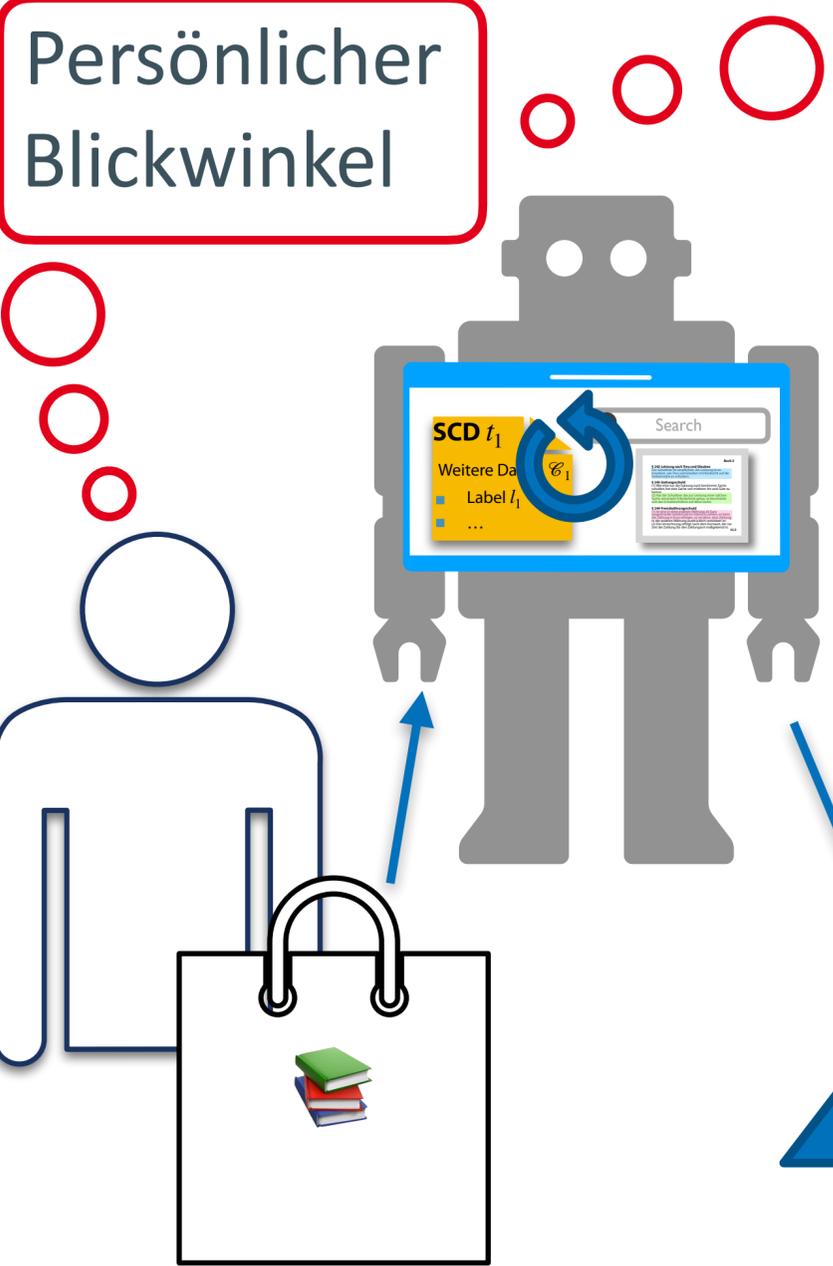
Verbessern

Verwenden

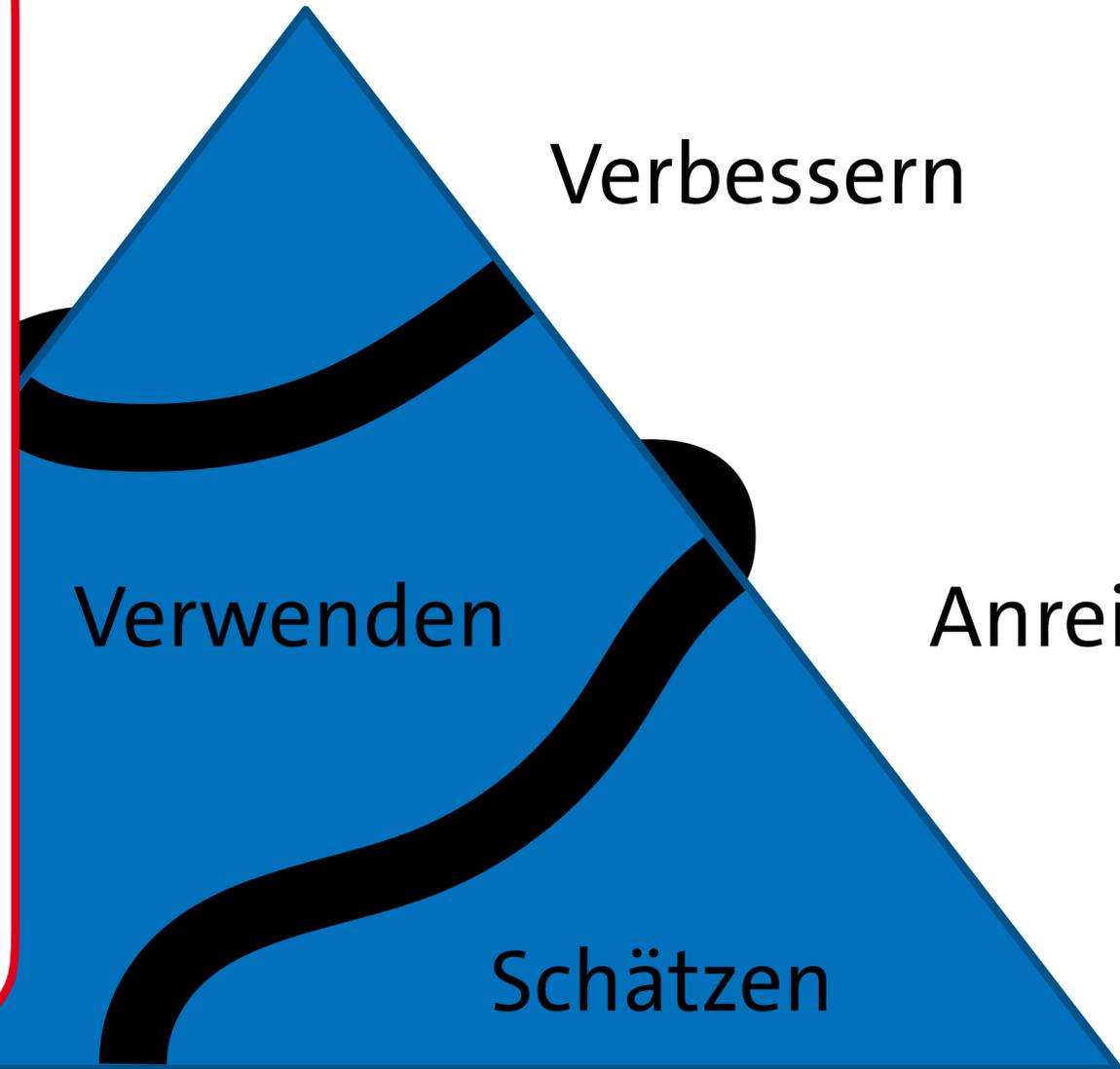
Schätzen

Anreichern

Persönlicher
Blickwinkel



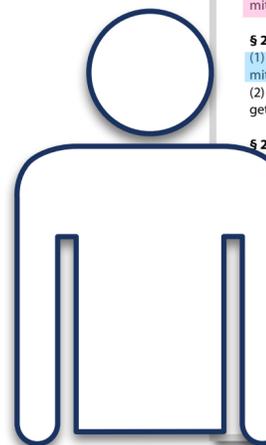
Erstellt mit *stable-diffusion-3-medium*



Blickwinkel

SCDs: Subjektive Inhaltsbeschreibungen

Blickwinkel



§ 242 Leistung nach Treu und Glauben **Buch 2**
 Der Schuldner ist verpflichtet, die Leistung so zu bewirken, wie Treu und Glauben mit Rücksicht auf die Verkehrssitte es erfordern.

§ 243 Gattungsschuld
 (1) Wer eine nur der Gattung nach bestimmte Sache schuldet, hat eine Sache von mittlerer Art und Güte zu leisten.
 (2) Hat der Schuldner das zur Leistung einer solchen Sache seinerseits Erforderliche getan, so beschränkt sich das Schuldverhältnis auf diese Sache.

§ 244 Fremdwährungsschuld
 Ist eine in einer anderen Währung als Euro ausgedrückte Geldschuld im Inland zu zahlen, so kann die Zahlung in Euro erfolgen, es sei denn, dass Zahlung in der anderen Währung ausdrücklich vereinbart ist.
 Die Umrechnung erfolgt nach dem Kurswert, der zur Zeit der Zahlung für den Zahlungsort maßgebend ist.

Geldsortenschuld
 Ist eine Geldschuld in einer bestimmten Münzsorte zu zahlen, die sich zur Zeit der Zahlung nicht mehr im Umlauf befindet, so ist die Zahlung so zu leisten, wie wenn die Münzsorte nicht bestimmt wäre.

Gesetzlicher Zinssatz
 Ist eine Schuld nach Gesetz oder Rechtsgeschäft zu verzinsen, so sind vier vom Hundert für das Jahr zu entrichten, sofern nicht ein anderes bestimmt ist.

Basiszinssatz * **BGB**

SCD t_2
 Weitere Daten \mathcal{C}_2
 ■ Label l_2
 ■ ...

SCD t_1
 Weitere Daten \mathcal{C}_1
 ■ Label l_1
 ■ ...

SCD t_3
 Weitere Daten \mathcal{C}_3
 ■ Label l_3
 ■ ...


 Search

§ 242 Leistung nach Treu und Glauben **Buch 2**
 Der Schuldner ist verpflichtet, die Leistung so zu bewirken, wie Treu und Glauben mit Rücksicht auf die Verkehrssitte es erfordern.

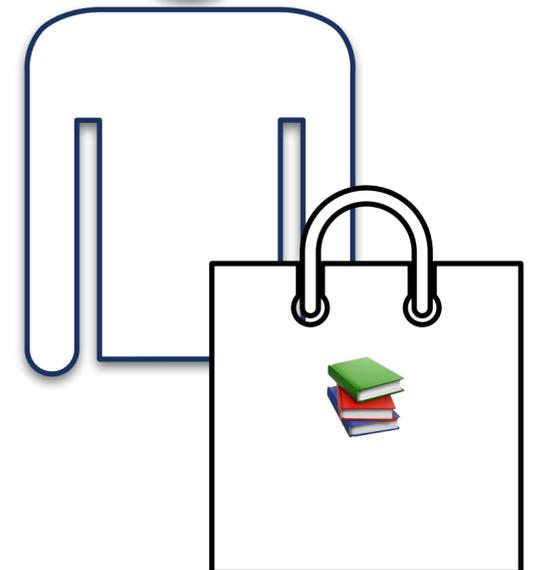
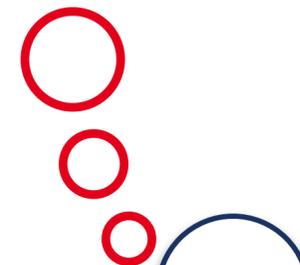
§ 243 Gattungsschuld
 (1) Wer eine nur der Gattung nach bestimmte Sache schuldet, hat eine Sache von mittlerer Art und Güte zu leisten.
 (2) Hat der Schuldner das zur Leistung einer solchen Sache seinerseits Erforderliche getan, so beschränkt sich das Schuldverhältnis auf diese Sache.

§ 244 Fremdwährungsschuld
 (1) Ist eine in einer anderen Währung als Euro ausgedrückte Geldschuld im Inland zu zahlen, so kann die Zahlung in Euro erfolgen, es sei denn, dass Zahlung in der anderen Währung ausdrücklich vereinbart ist.
 (2) Die Umrechnung erfolgt nach dem Kurswert, der zur Zeit der Zahlung für den Zahlungsort maßgebend ist.

§ 245 Geldsortenschuld
 Ist eine Geldschuld in einer bestimmten Münzsorte zu zahlen, die sich zur Zeit der Zahlung nicht mehr im Umlauf befindet, so ist die Zahlung so zu leisten, wie wenn die Münzsorte nicht bestimmt wäre.

§ 246 Gesetzlicher Zinssatz
 Ist eine Schuld nach Gesetz oder Rechtsgeschäft zu verzinsen, so sind vier vom Hundert für das Jahr zu entrichten, sofern nicht ein anderes bestimmt ist.

§ 247 Basiszinssatz * **BGB**



- Bisher: Fokus auf Konzeption und Verwendung vorhandener SCDs
- Offen: Erstellen von initialen SCDs (*schätzen*), *anreichern* und *verbessern*

Schätzen: USEM – UnSupervised Estimator of SCD Matrices

$$\begin{matrix}
 & w_1 & w_2 & \dots & w_L \\
 t_1 & v_{1,1} & v_{1,2} & \dots & v_{1,L} \\
 t_2 & v_{2,1} & v_{2,2} & \dots & v_{2,L} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 t_K & v_{K,1} & v_{K,2} & \dots & v_{K,L}
 \end{matrix}$$

Buch 2

§ 242 Leistung nach Treu und Glauben
 Der Schuldner ist verpflichtet, die Leistung so zu bewirken, wie Treu und Glauben mit Rücksicht auf die Verkehrssitte es erfordern.

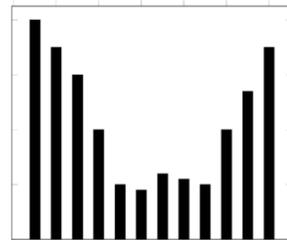
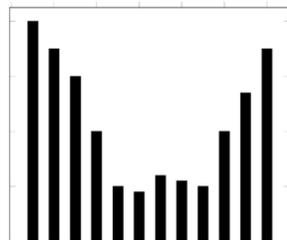
§ 243 Gattungsschuld
 (1) Wer eine nur der Gattung nach bestimmte Sache schuldet, hat eine Sache von mittlerer Art und Güte zu leisten.
 (2) Hat der Schuldner das zur Leistung einer solchen Sache seinerseits Erforderliche getan, so beschränkt sich das Schuldverhältnis auf diese Sache.

§ 244 Fremdwährungsschuld
 (1) Ist eine in einer anderen Währung als Euro ausgedrückte Geldschuld im Inland zu zahlen, so kann die Zahlung in Euro erfolgen, es sei denn, dass Zahlung in der anderen Währung ausdrücklich vereinbart ist.
 (2) Die Umrechnung erfolgt nach dem Kurswert, der zur Zeit der Zahlung für den Zahlungsort maßgebend ist.

§ 245 Geldsortenschuld
 Ist eine Geldschuld in einer bestimmten Münzsorte zu zahlen, die sich zur Zeit der Zahlung nicht mehr im Umlauf befindet, so ist die Zahlung so zu leisten, wie wenn die Münzsorte nicht bestimmt wäre.

§ 246 Gesetzlicher Zinssatz
 Ist eine Schuld nach Gesetz oder Rechtsgeschäft zu verzinsen, so sind vier vom Hundert für das Jahr zu entrichten, sofern nicht ein anderes bestimmt ist.

§ 247 Basiszinssatz* BGB

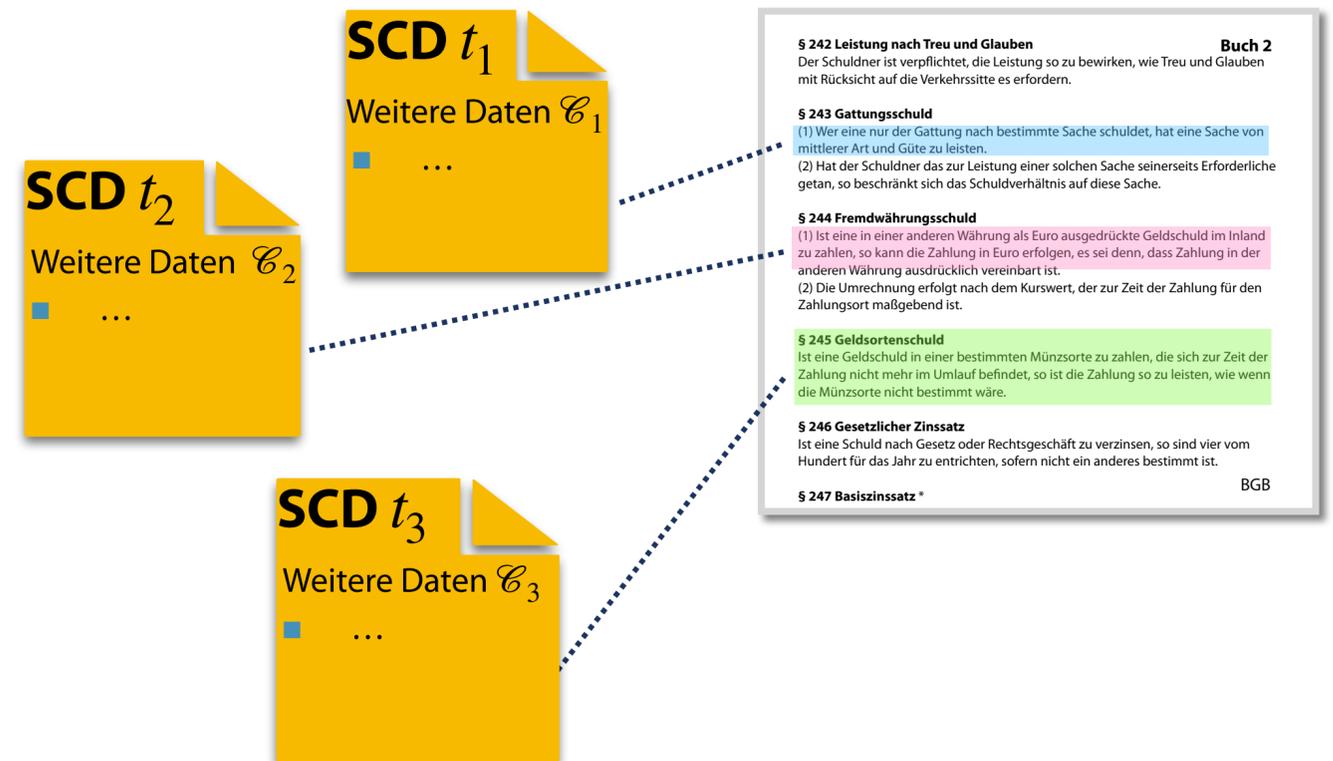


$$d = (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}, \dots, w_N)$$

Schätzen: USEM – UnSupervised Estimator of SCD Matrices

- Korpus ohne jegliche SCDs
- Ziel: Bestimme initiale Menge von SCDs

1. Eine SCD pro Satz
2. Identifizieren von *ähnlichen* SCDs (Gierig, K-Means, DBSCAN)
3. Zusammenfassen von *ähnlichen* SCDs
4. *Bestes* Modell wählen

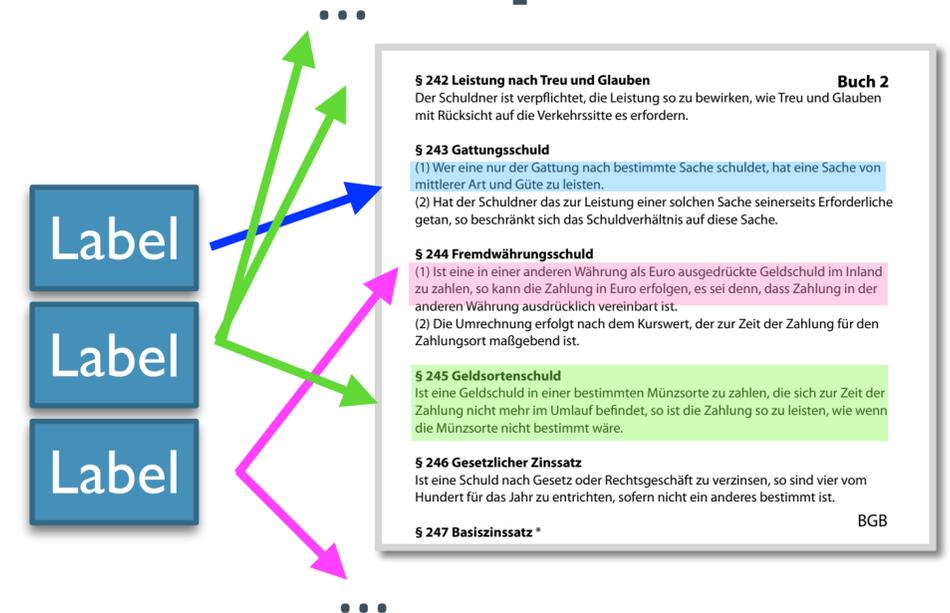


Anreichern: LESS – Label Estimation for SCDs without Supervision

- Korpus mit *initialen* SCDs
- Ziel: SCDs Menschen präsentieren
 - IR-System bekommt Anfrage (Wörter, Sätze) von Menschen
 - IR-System beantwortet mit passenden SCDs (& Dokumenten)

Ergebnisse:

1. SCD t_1
2. SCD t_2
3. SCD t_3



LESS: Informationsquellen

- Verfügbar pro SCD

- ✓ Referenzierte Sätze $\{s_1, \dots, s_S\}$

- ✓ Wortverteilung $\{v_{i,1}, \dots, v_{i,L}\}$

- ✗ Label

- ✗ Sonstige Daten wie z.B. *Relationen*

- Formalisiertes Problem

$$l_i = \underset{l_j \in \text{alle möglichen Labels}}{\operatorname{argmax}} \operatorname{Utility}\left(l_j, t_i = \left((v_{i,1}, \dots, v_{i,L}), \{s_1, \dots, s_S\}\right)\right)$$

SCD t_i

Weitere Daten \mathcal{C}_i :

- Label l_i
- Relationen
- Links
- ...

Referenzierte
Sätze

$\{s_1, \dots, s_S\}$



Wortverteilung
 $\{v_{i,1}, \dots, v_{i,L}\}$

$$\begin{matrix} & w_1 & w_2 & \dots & w_L \\ t_1 & v_{1,1} & v_{1,2} & \dots & v_{1,L} \\ t_2 & v_{2,1} & v_{2,2} & \dots & v_{2,L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_K & v_{K,1} & v_{K,2} & \dots & v_{K,L} \end{matrix}$$

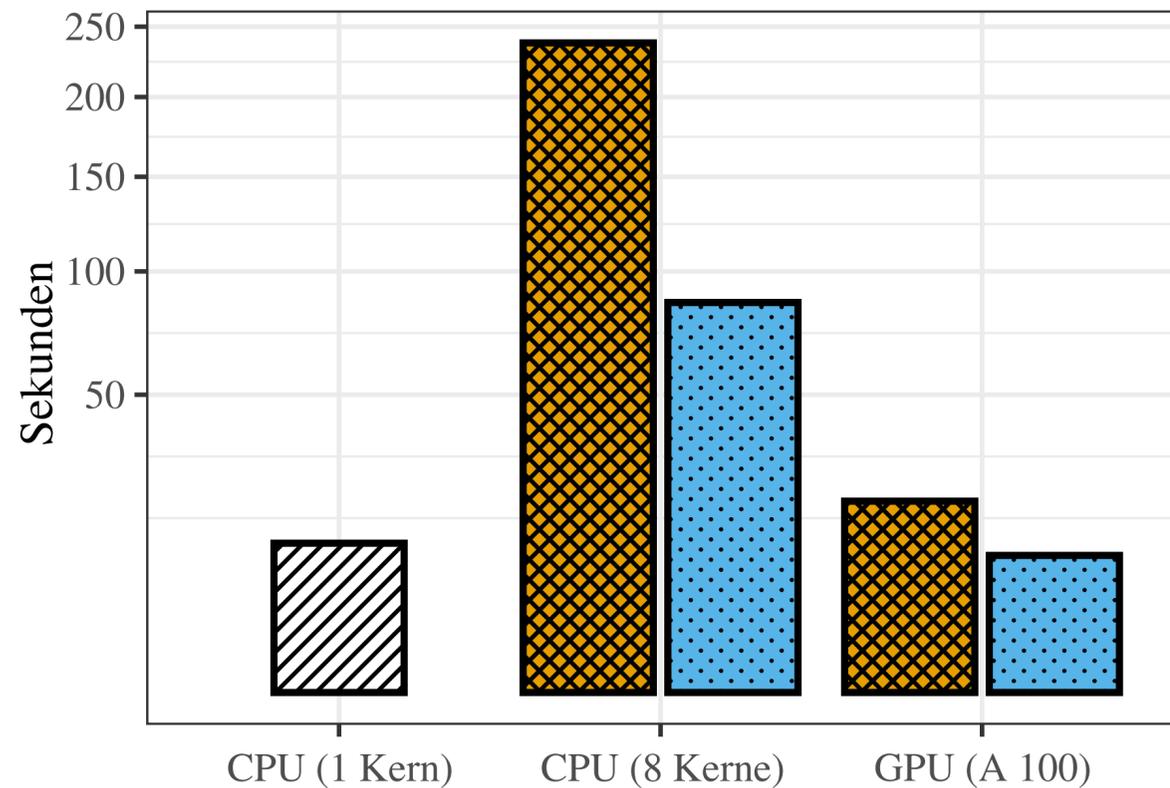
LESS: Kandidaten für und Nützlichkeit von Labeln

- Wie sollte ein Label aussehen?
 - Sequenz von Wörtern, eine kurze Beschreibung
 - Zusammenfassung einer SCD
- **Kandidaten:**
 - Referenzierte Sätze $\{s_1, \dots, s_S\}$
 - Umformuliertes Problem
- Was macht ein *gutes* Label aus?
 - Ähnlich zu referenzierten Sätzen einer SCD
 - Wortverteilung generiert Sätze
- **Nützlichkeit:** Kosinusähnlichkeit
 - Nutze Wortverteilung $\{v_{i,1}, \dots, v_{i,L}\}$
 - Umformuliertes Problem

$$l_i = \arg \max_{s_j \in \{s_1, \dots, s_S\}} Utility(s_j, (v_{i,1}, \dots, v_{i,L}))$$

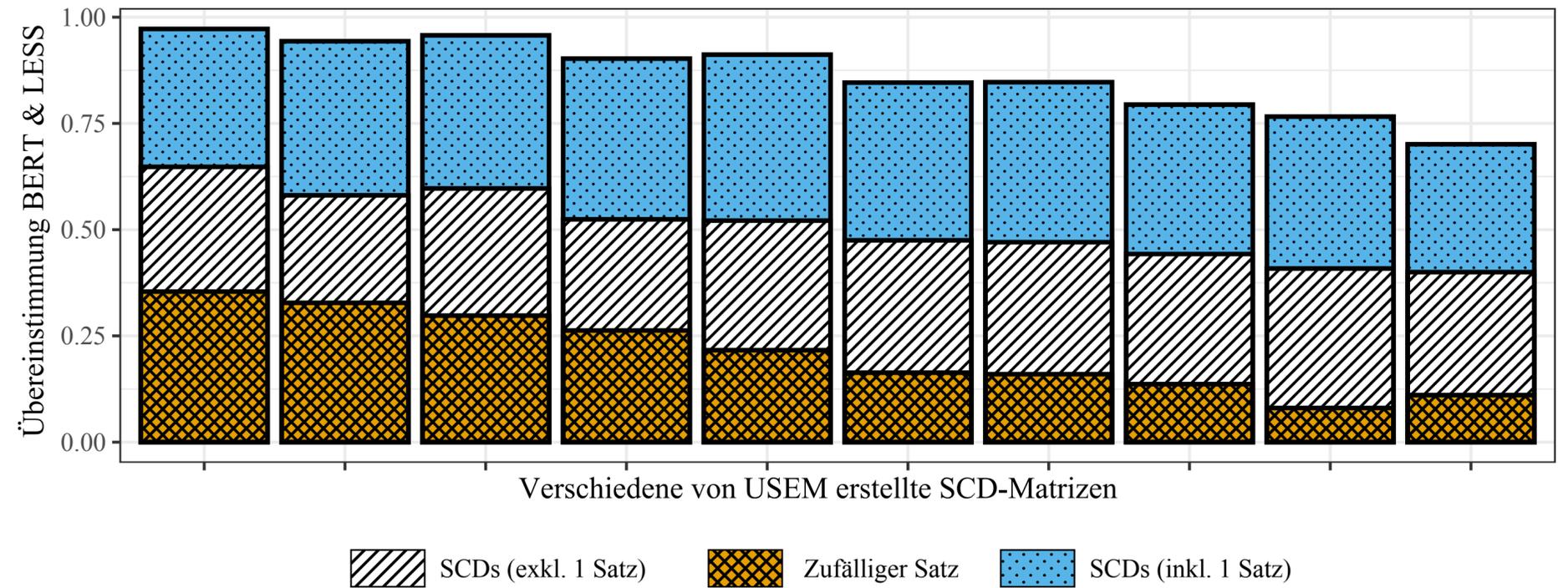
$$l_i = \arg \max_{s_j \in \{s_1, \dots, s_S\}} \frac{\vec{s}_j \cdot v_i}{\|\vec{s}_j\|_2 \cdot \|v_i\|_2}$$

LESS: Evaluation



 LESS
  BERT Q&A
  BERT Vektoren

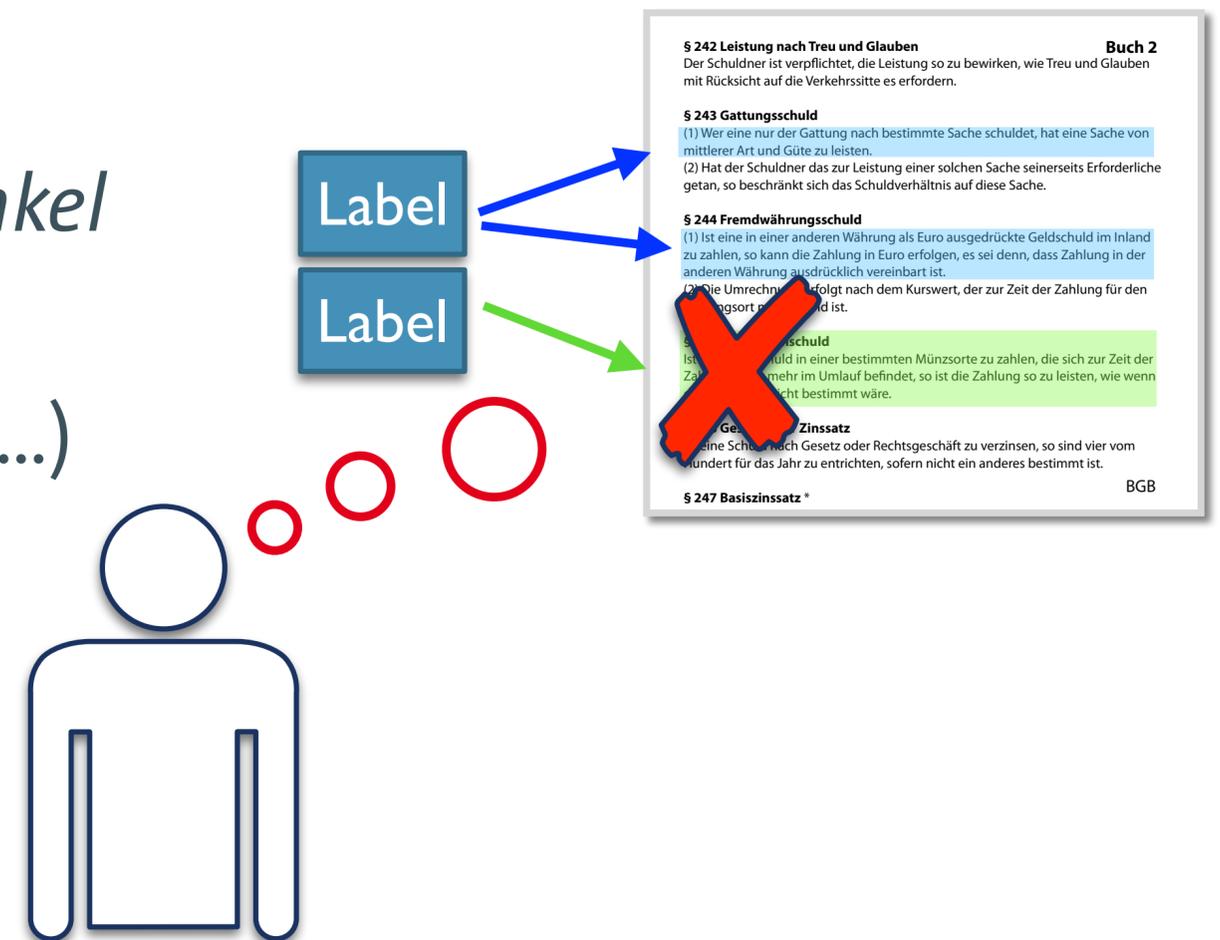
Zwei Techniken nutzen BERT; 1 oder 8 Intel CPU-Kerne und eine einzelne NVIDIA A100 40GB GPU



Zufälliger Satz: Theoretische Übereinstimmung, die ein zufällig wählender Ansatz erreichen würde.

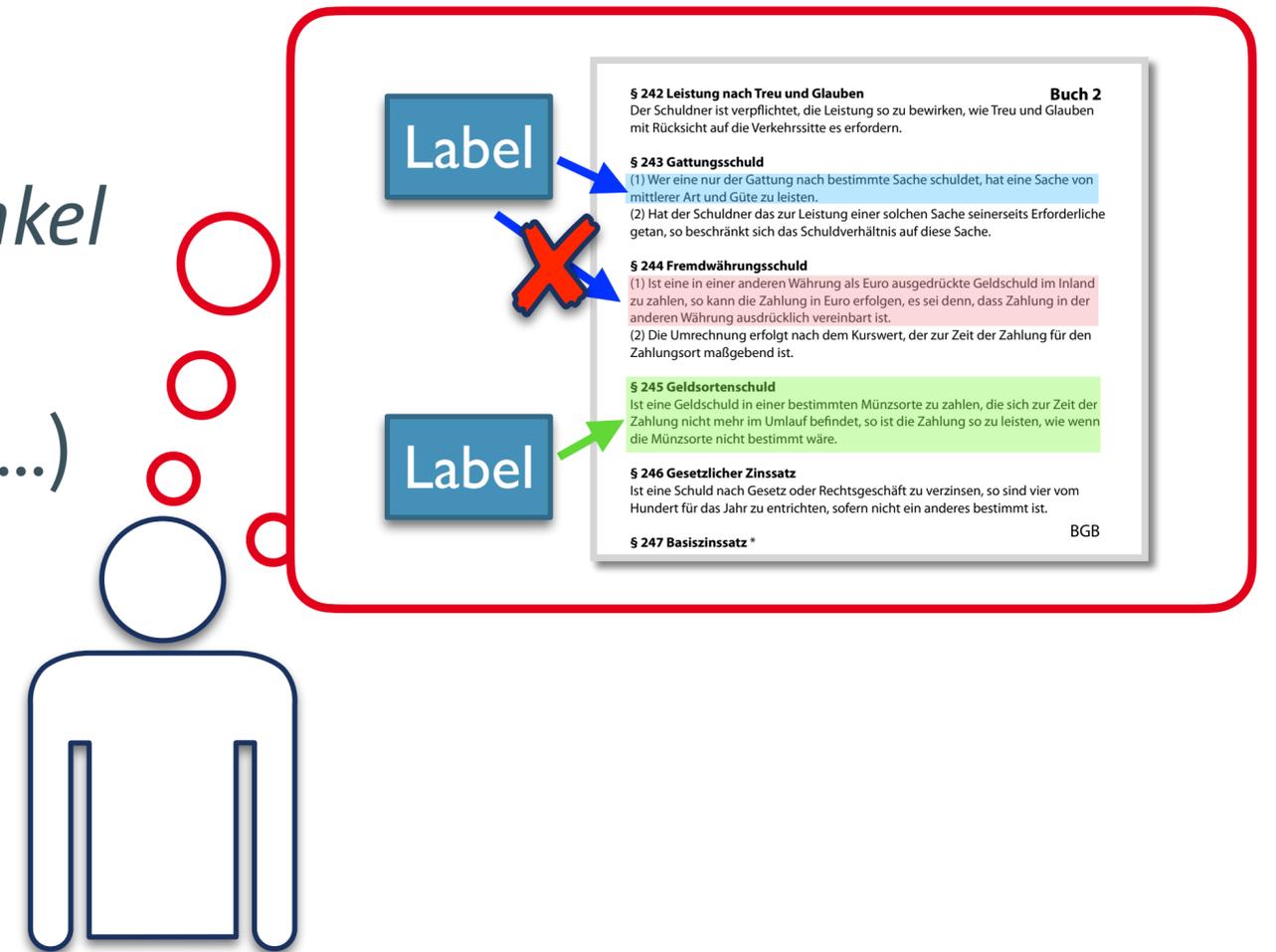
Verbessern: FrESH & ReFrESH – (Relation-preserving) Feedback-reliant Enh. of SCDs

- Korpus mit SCDs inkl. Label
- Problem: SCDs oder Text passen nicht zum *Blickwinkel*
 1. **FrESH**: Text im Korpus ist falsch und soll entfernt werden (Fake News, Urheberrecht, Datenschutz, ...)



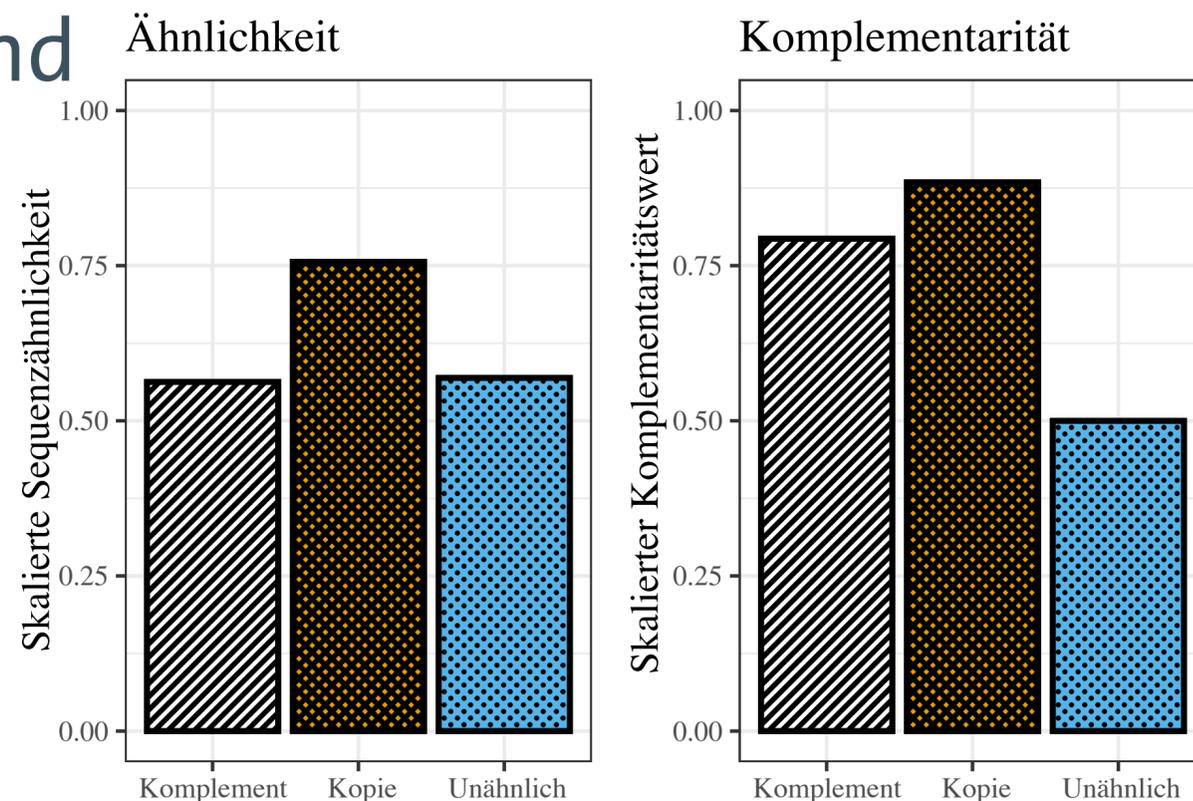
Verbessern: FrESH & ReFrESH – (Relation-preserving) Feedback-reliant Enh. of SCDs

- Korpus mit SCDs inkl. Label
- Problem: SCDs oder Text passen nicht zum *Blickwinkel*
 1. **FrESH**: Text im Korpus ist falsch und soll entfernt werden (Fake News, Urheberrecht, Datenschutz, ...)
 2. **ReFrESH**: SCD passt nicht zu referenziertem Satz
- Inkrementelle Änderung an den SCDs
- Potentiell zuvor Änderungen durch Menschen
→ keine Invertierung möglich
- Berücksichtigung von Relationen und weiteren Daten



Verwenden: Komplementäre SCDs

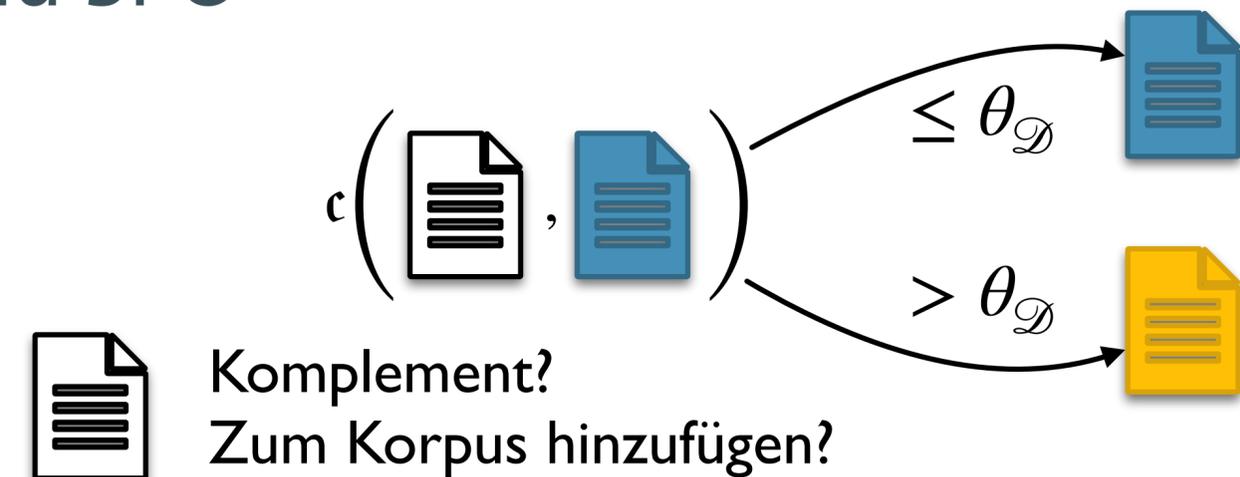
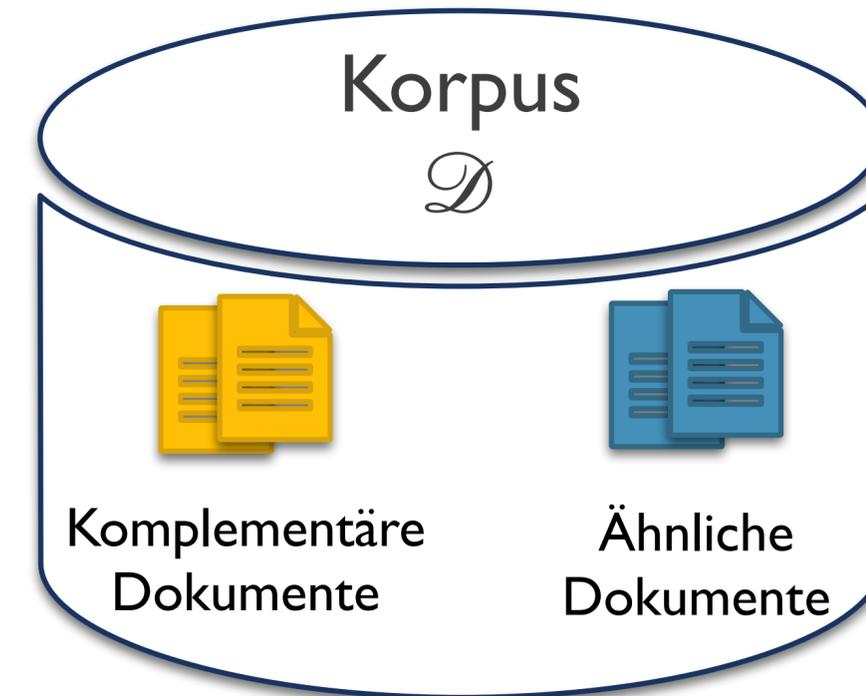
- Korpus mit SCDs inkl. Label
- Relationen zwischen SCDs am Beispiel Komplementarität
- Ziel: Klassifikation von Dokumenten, die *komplementär* zum Korpus sind

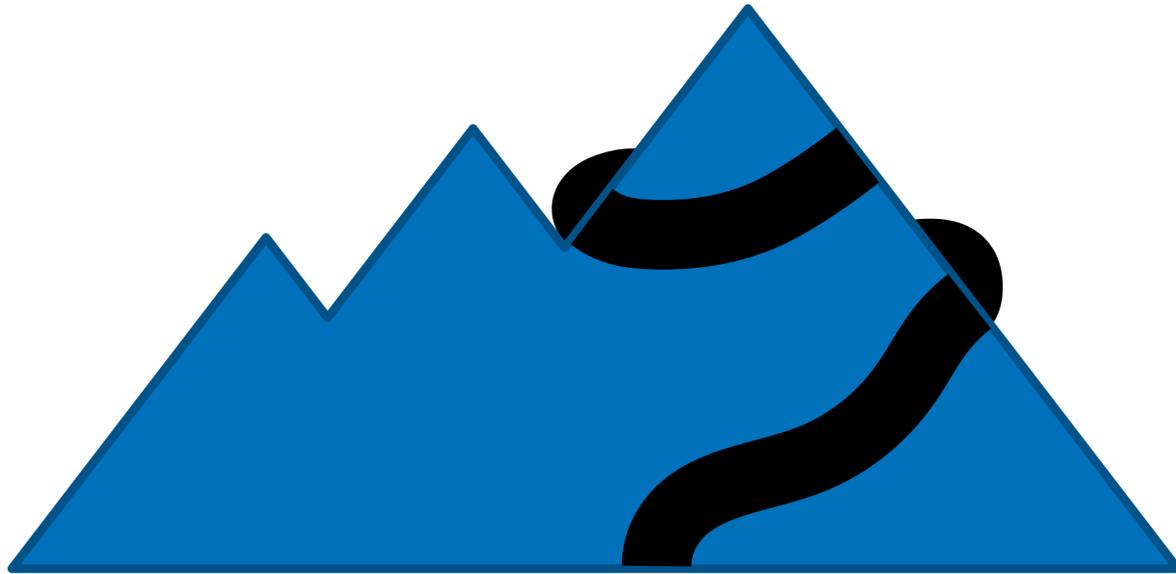


Komplement?
Zum Korpus hinzufügen?

Verwenden: Komplementäre SCDs

- Korpus mit SCDs inkl. Label
- Relationen zwischen SCDs am Beispiel Komplementarität
- Ziel: Klassifikation von Dokumenten, die *komplementär* zum Korpus sind
- Formelle Definition auf Basis einer Taxonomie und SPO-Tripeln
- Erweiterter Korpus, SCD-Matrix und angepasste Algorithmen zur Dokumentenklassifikation



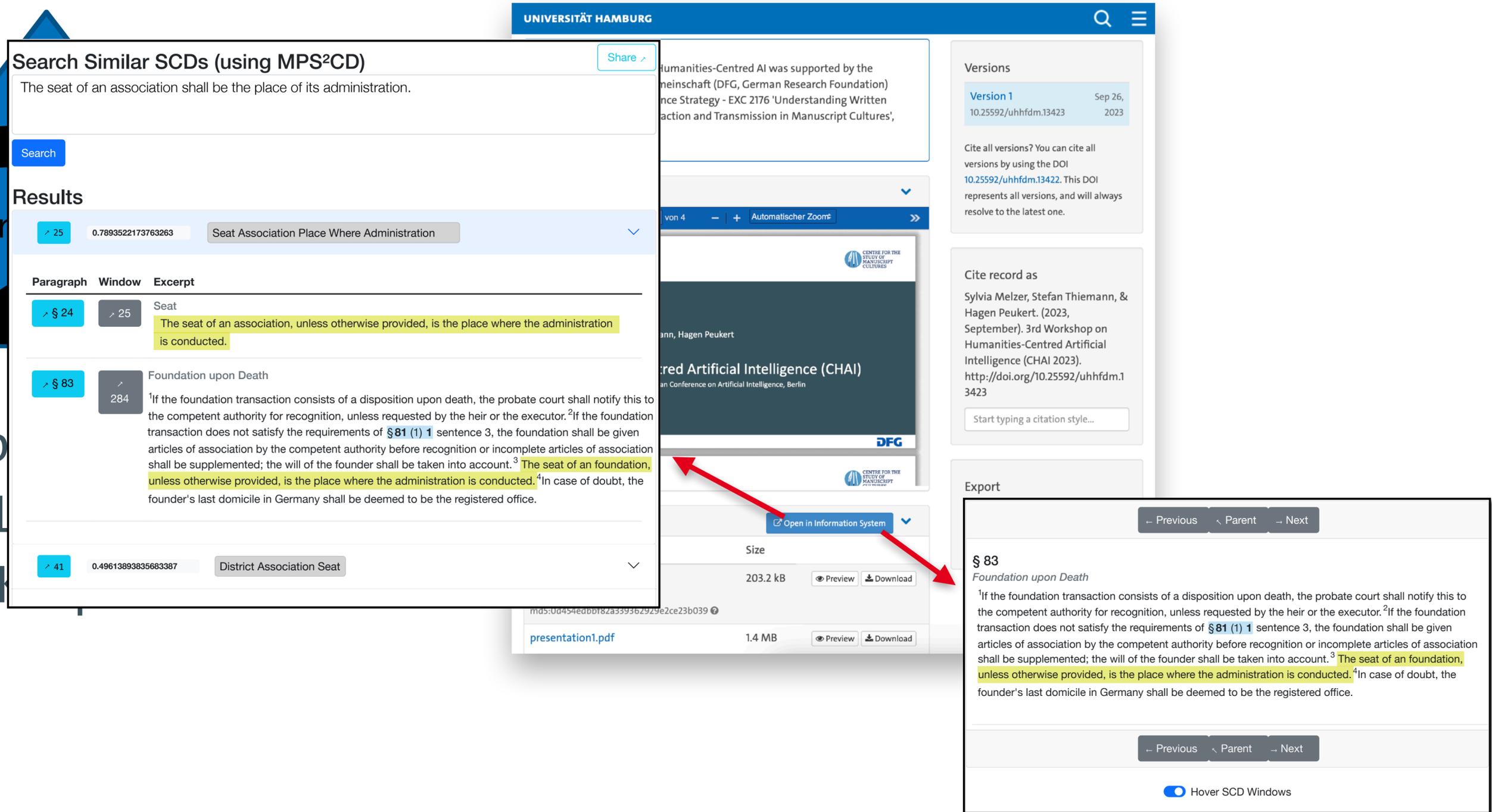


- ✓ Automatisiert persönliches SCD-Modell (USEM, LESS, (Re-)FrESH) für mitgebrachten Korpus erstellen

Rückblick:

- Bisher: Fokus auf Konzeption und Verwendung vorhandener SCDs
- Offen: Erstellen von initialen SCDs (*schätzen*), *anreichern* und *verbessern*

Verwer
 Schätzen
 ✓ Automatisiert p
 Modell (USEM, L
 mitgebrachten k



Search Similar SCDs (using MPS²CD) Share >
 The seat of an association shall be the place of its administration.

Search

Results

Paragraph	Window	Excerpt
> 25	0.7893522173763263	Seat Association Place Where Administration
> § 24	> 25	Seat The seat of an association, unless otherwise provided, is the place where the administration is conducted.
> § 83	> 284	Foundation upon Death ¹ If the foundation transaction consists of a disposition upon death, the probate court shall notify this to the competent authority for recognition, unless requested by the heir or the executor. ² If the foundation transaction does not satisfy the requirements of § 81 (1) 1 sentence 3, the foundation shall be given articles of association by the competent authority before recognition or incomplete articles of association shall be supplemented; the will of the founder shall be taken into account. ³ The seat of an foundation, unless otherwise provided, is the place where the administration is conducted. ⁴ In case of doubt, the founder's last domicile in Germany shall be deemed to be the registered office.
> 41	0.49613893835683387	District Association Seat

Humanities-Centred AI was supported by the
 meinschaft (DFG, German Research Foundation)
 nce Strategy - EXC 2176 'Understanding Written
 action and Transmission in Manuscript Cultures',

von 4 - + Automatischer Zoom >>

ann, Hagen Peukert
 red Artificial Intelligence (CHAI)
 an Conference on Artificial Intelligence, Berlin

Versions
Version 1 Sep 26, 2023
 10.25592/uhhfdm.13423

Cite all versions? You can cite all versions by using the DOI 10.25592/uhhfdm.13422. This DOI represents all versions, and will always resolve to the latest one.

Cite record as
 Sylvia Melzer, Stefan Thiemann, & Hagen Peukert. (2023, September). 3rd Workshop on Humanities-Centred Artificial Intelligence (CHAI 2023). <http://doi.org/10.25592/uhhfdm.13423>
 Start typing a citation style...

Export

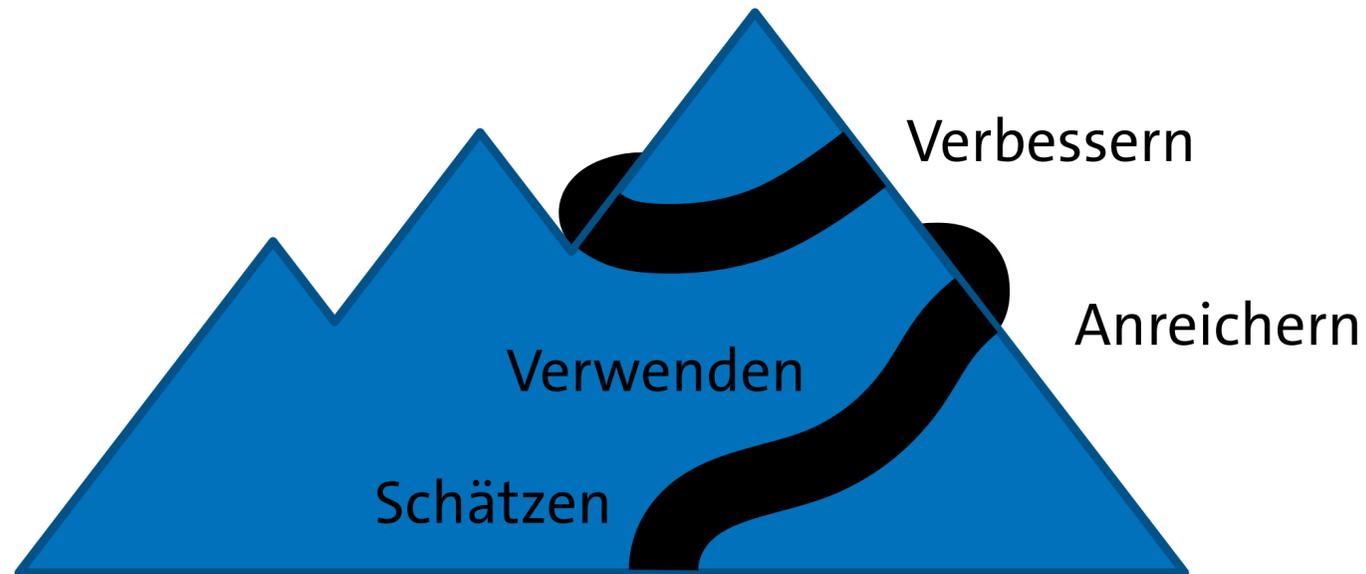
Open in Information System

Size
 203.2 kB Preview Download

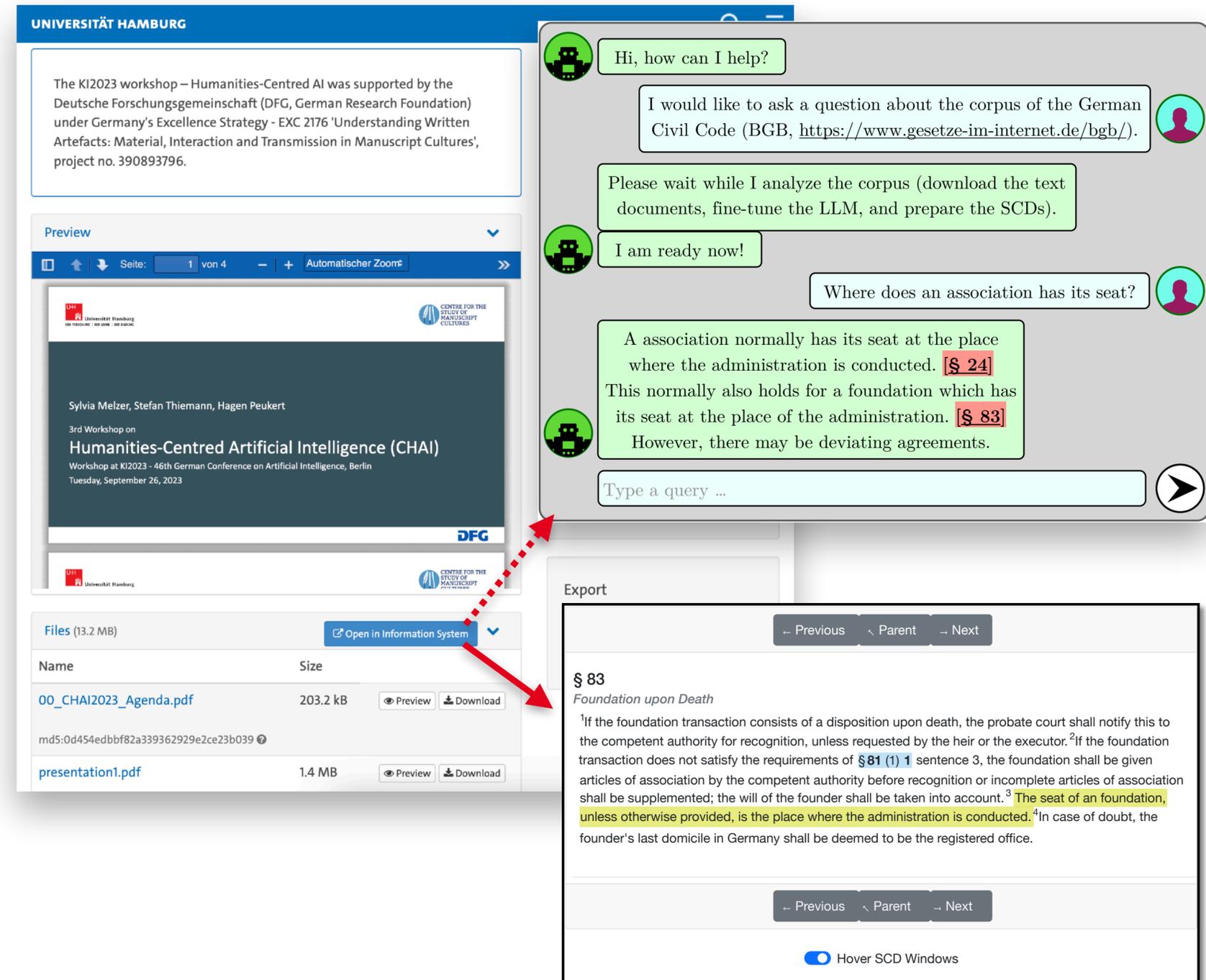
presentation1.pdf 1.4 MB Preview Download

§ 83
Foundation upon Death
¹If the foundation transaction consists of a disposition upon death, the probate court shall notify this to the competent authority for recognition, unless requested by the heir or the executor. ²If the foundation transaction does not satisfy the requirements of § 81 (1) 1 sentence 3, the foundation shall be given articles of association by the competent authority before recognition or incomplete articles of association shall be supplemented; the will of the founder shall be taken into account. ³The seat of an foundation, unless otherwise provided, is the place where the administration is conducted. ⁴In case of doubt, the founder's last domicile in Germany shall be deemed to be the registered office.

Hover SCD Windows



Automatisiert persönliches SCD-Modell (USEM, LESS, (Re-)FrESH) für mitgebrachten Korpus erstellen



UNIVERSITÄT HAMBURG

The KI2023 workshop – Humanities-Centred AI was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796.

Preview

Seite: 1 von 4 Automatischer Zoom

Sylvia Melzer, Stefan Thiemann, Hagen Peukert
 3rd Workshop on
Humanities-Centred Artificial Intelligence (CHAI)
 Workshop at KI2023 - 46th German Conference on Artificial Intelligence, Berlin
 Tuesday, September 26, 2023

Files (13.2 MB) Open in Information System

Name	Size	Preview	Download
00_CHAI2023_Agenda.pdf	203.2 kB		
md5:0d454edbbf82a339362929e2ce23b039			
presentation1.pdf	1.4 MB		

Hi, how can I help?

I would like to ask a question about the corpus of the German Civil Code (BGB, <https://www.gesetze-im-internet.de/bgb/>).

Please wait while I analyze the corpus (download the text documents, fine-tune the LLM, and prepare the SCDs).

I am ready now!

Where does an association has its seat?

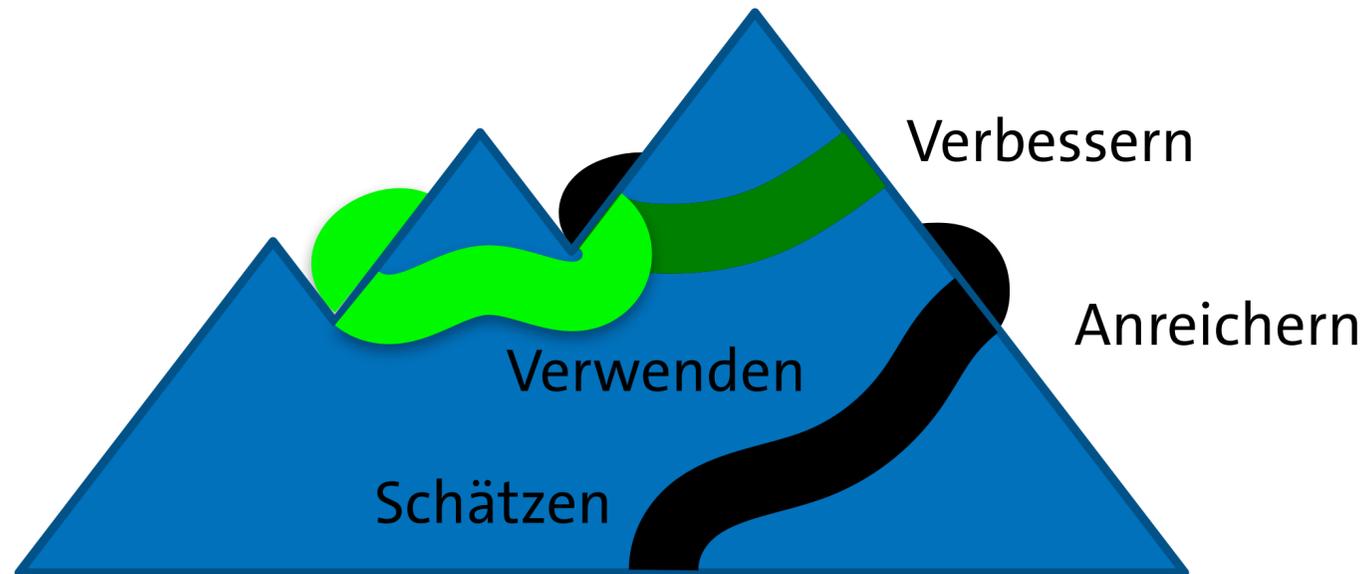
A association normally has its seat at the place where the administration is conducted. **§ 24**
 This normally also holds for a foundation which has its seat at the place of the administration. **§ 83**
 However, there may be deviating agreements.

Type a query ...

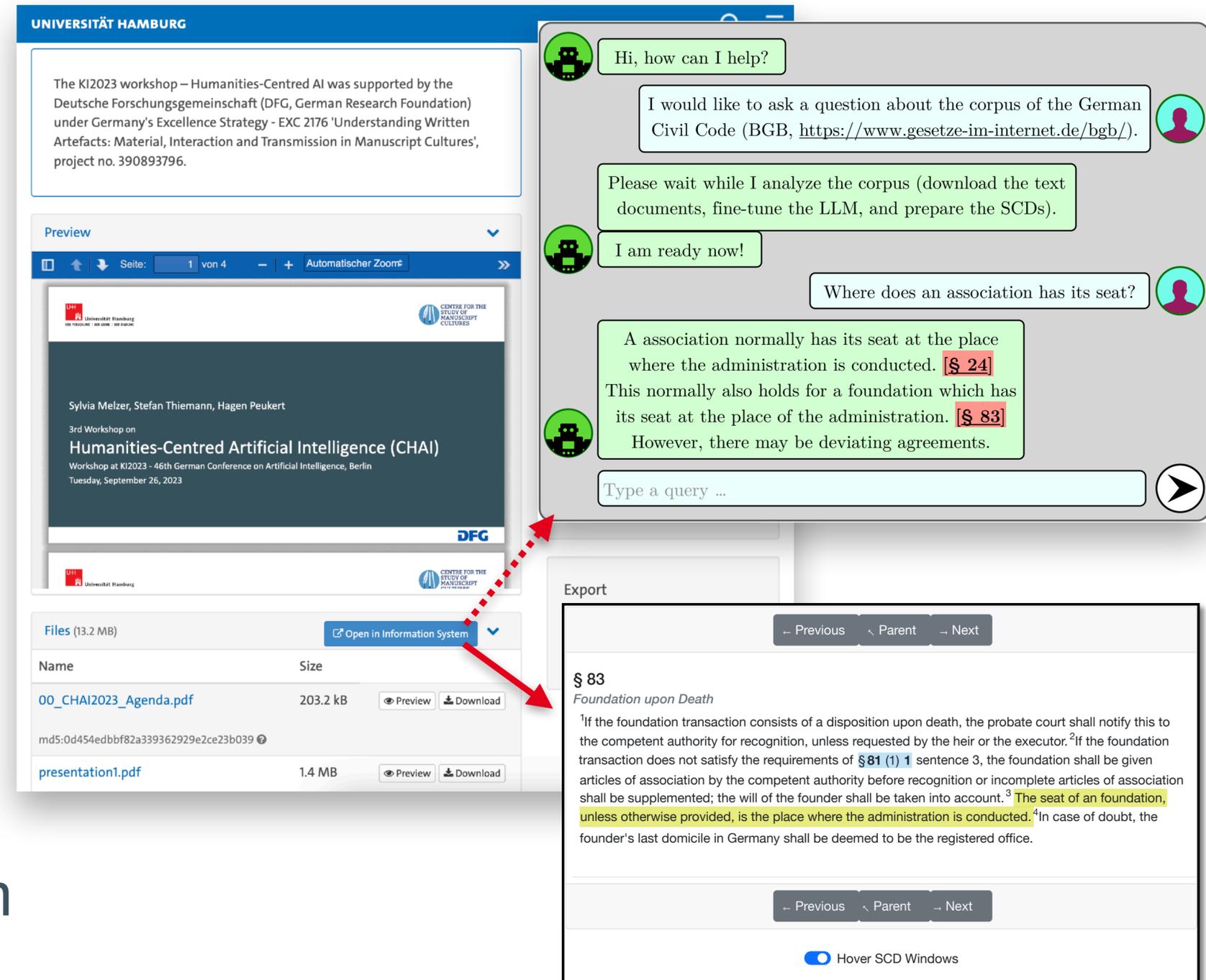
Export

§ 83
Foundation upon Death
¹If the foundation transaction consists of a disposition upon death, the probate court shall notify this to the competent authority for recognition, unless requested by the heir or the executor. ²If the foundation transaction does not satisfy the requirements of **§ 81 (1) 1** sentence 3, the foundation shall be given articles of association by the competent authority before recognition or incomplete articles of association shall be supplemented; the will of the founder shall be taken into account. ³The seat of an foundation, unless otherwise provided, is the place where the administration is conducted. ⁴In case of doubt, the founder's last domicile in Germany shall be deemed to be the registered office.

Hover SCD Windows



- Automatisiert persönliches SCD-Modell (USEM, LESS, (Re-)FrESH) für mitgebrachten Korpus erstellen
- Modell auf einen neuen Korpus übertragen und Modelle kombinieren



The screenshot displays the CHAI interface. On the left, a document preview shows a workshop agenda for 'Humanities-Centred Artificial Intelligence (CHAI)' on September 26, 2023. Below the preview is a file list with two items: '00_CHAI2023_Agenda.pdf' (203.2 kB) and 'presentation1.pdf' (1.4 MB). On the right, a chat window shows a conversation with an AI assistant. The assistant asks 'Hi, how can I help?' and responds to a query about the German Civil Code (BGB) by stating it is ready to analyze the corpus. A subsequent query asks 'Where does an association has its seat?', and the assistant provides a detailed answer citing legal sections § 24 and § 83. Below the chat, a detailed SCD view for § 83 is shown, titled 'Foundation upon Death', with a 'Hover SCD Windows' toggle.

Literaturverzeichnis

- Asselborn et al. (CHAI 2023)
Thomas Asselborn, Sylvia Melzer, Said Aljoumani, Magnus Bender, Florian Andreas Marwitz, Konrad Hirschler, Ralf Möller: **Fine-tuning BERT Models on Demand for Information Systems Explained Using Training Data from Pre-modern Arabic** in *Humanities Centred AI (CHAI) Workshop at KI 2023*
- Asselborn et al. (HSSC 202?)
Thomas Asselborn, Sylvia Melzer, Simon Schiff, Said Aljoumani, Magnus Bender, Florian Andreas Marwitz, Konrad Hirschler and Ralf Möller: **Building Sustainable Information Systems and Transformer Models on Demand** eingereicht für *Humanities and Social Sciences Communications, Nature*
- B. et al. (ICSC 2022)
Magnus Bender, Felix Kuhr, Tanya Braun: **To Extend or not to Extend: Complementary Documents** in *IEEE International Conference on Semantic Computing, 2022*
- B. et al. (IJSC 2022)
Magnus Bender, Felix Kuhr, Tanya Braun: **To Extend or not to Extend? Enriching a Corpus with Complementary and Related Documents** in *International Journal of Semantic Computing, 2022*
- B. et al. (ICSC 2023)
Magnus Bender, Tanya Braun, Ralf Möller, Marcel Gehrke: **Unsupervised Estimation of Subjective Content Descriptions** in *IEEE International Conference on Semantic Computing, 2023*
- B. et al. (CHAI 2023)
Magnus Bender, Kira Schwandt, Ralf Möller, Marcel Gehrke: **FrESH – Feedback-reliant Enhancement of Subjective Content Descriptions by Humans** in *Humanities Centred AI (CHAI) Workshop at KI 2023*
- B. et al. (KI 2023)
Magnus Bender, Tanya Braun, Ralf Möller, Marcel Gehrke: **LESS is More LEan Computing for Selective Summaries** in *KI 2023: Advances in Artificial Intelligence, Springer*
- B. et al. (IJSC 2024)
Magnus Bender, Tanya Braun, Ralf Möller, Marcel Gehrke: **Unsupervised Estimation of Subjective Content Descriptions in Information System** in *International Journal of Semantic Computing, 2024*
- B. et al. (ICSC 2024)
Magnus Bender, Tanya Braun, Ralf Möller, Marcel Gehrke: **ReFrESH – Relation-preserving Feedback-reliant Enhancement of Subjective Content Descriptions** in *IEEE International Conference on Semantic Computing, 2024 – Best Paper Award*
- B. et al. (IJSC 202?)
Magnus Bender, Tanya Braun, Ralf Möller, Marcel Gehrke: **Enhancement of Subjective Content Descriptions by using Human Feedback** wird veröffentlicht in *International Journal of Semantic Computing* (entstanden nach der Dissertation)
- Devlin et al. (2019)
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding** in *North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019*
- Ester et al. (1996)
Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu: **A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise** in *International Conference on Knowledge Discovery and Data Mining, 1996*
- Kuhr et al. (AI 2019)
Felix Kuhr, Tanya Braun, Magnus Bender, Ralf Möller: **To Extend or not to Extend? Context-specific Corpus Enrichment** in *Proceedings of AI 2019: Advances in Artificial Intelligence, Springer*
- Kuhr et al. (IJSC 2020)
Felix Kuhr, Magnus Bender, Tanya Braun, Ralf Möller: **Augmenting and Automating Corpus Enrichment** in *International Journal of Semantic Computing, 2020*
- Lloyd (1982)
Stuart P. Lloyd: **Least squares quantization in PCM** in *IEEE Transactions on Information Theory, 1982*