



# Kollaborative Lexikographie

Strukturen, Dynamik und Zusammensetzung  
gemeinschaftlich erarbeiteter Wortschätze

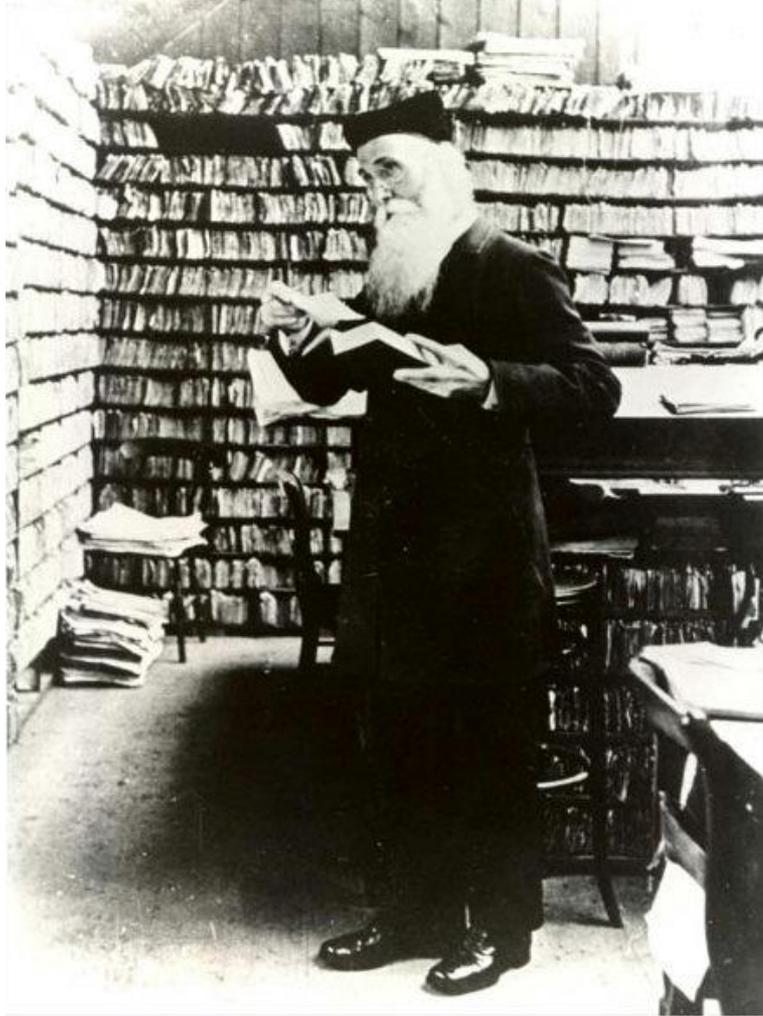
**Christian M. Meyer**

Technische Universität Darmstadt

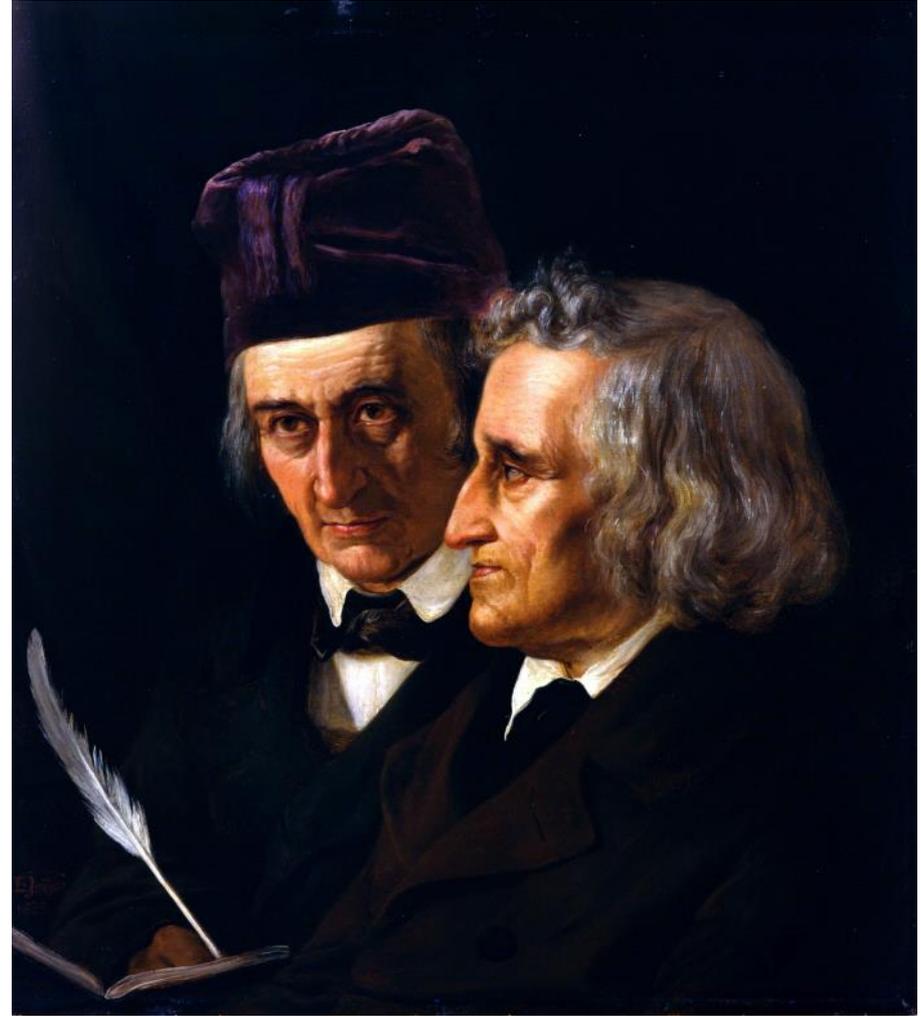
Ubiquitous Knowledge Processing Lab

53. Jahrestagung des Instituts für Deutsche Sprache „Wortschätze:  
Dynamik, Muster, Komplexität“, Mannheim, 14.–16. März 2017.

# Das traditionelle Bild der Lexikographie



James Murray, editor and philologist (before 1910)  
(Public domain, <https://commons.wikimedia.org/wiki/File:James-Murray.jpg>)



Doppelporträt der Brüder Wilhelm Grimm (links) und Jacob Grimm (rechts) von Elisabeth Jerichau-Baumann, 1855  
(Gemeinfrei, [https://de.wikipedia.org/wiki/B%C3%BCder\\_Grimm#/media/File:Grimm.jpg](https://de.wikipedia.org/wiki/B%C3%BCder_Grimm#/media/File:Grimm.jpg))

# Das traditionelle Bild der Lexikographie

Die Erarbeitung eines Wörterbuchs durch...

- einzelne Autoren oder überschaubare Teams
- mit klaren Richtlinien zur Ausarbeitung der Artikel
- als zentrales Produkt eines Verlags oder (wissenschaftlichen) Instituts

**Akademische und  
Verlagslexikographie**

# Neuer Ansatz: Kollaborative Lexikographie

Die Erarbeitung eines Wörterbuchs durch...

- einzelne Autoren oder überschaubare Teams
- mit klaren Richtlinien zur Ausarbeitung der Artikel
- als zentrales Produkt eines Verlags oder (wissenschaftlichen) Instituts

**Akademische und  
Verlagslexikographie**

- eine große Anzahl Freiwilliger („Nutzer“)

**Kollaborative  
Lexikographie**

# Neuer Ansatz: Kollaborative Lexikographie

Die Erarbeitung eines Wörterbuchs durch...

- einzelne Autoren oder überschaubare Teams
- mit klaren Richtlinien zur Ausarbeitung der Artikel
- als zentrales Produkt eines Verlags oder (wissenschaftlichen) Instituts

**Akademische und  
Verlagslexikographie**

- eine große Anzahl Freiwilliger („Nutzer“)
- evolutionär durch Aushandlung und Diskussion
- als frei lizenziertes Wissen oder Ergänzung bestehender Wörterbücher

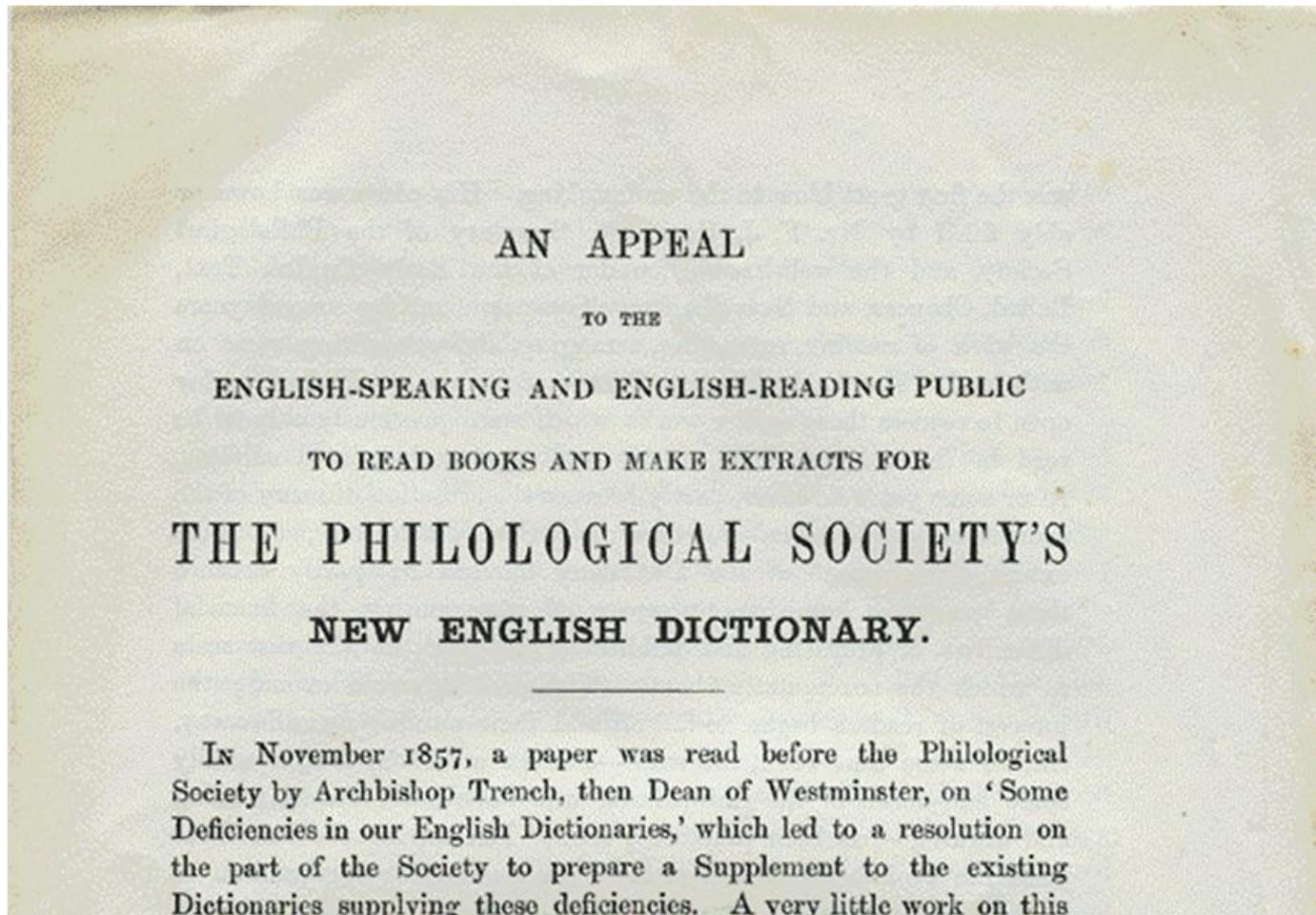
**Kollaborative  
Lexikographie**

# Nutzerbeteiligung ist kein neues Thema



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

April 1879:



An appeal to the English-speaking and English-reading public to read books and make extracts for the Philological Society's New English Dictionary (April 1879), <http://public.oed.com/wpcms/wp-content/uploads/Appeal047911.gif>

(Thier, 2014)

# Aber im Web 2.0 ein Massenphänomen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



# Nutzerbeteiligung wird enorm vielfältig

## Direkte Beteiligung

Beiträge zu

- **offen-kollaborativen,**
- **kollaborativ-  
institutionellen,**
- **semi-kollaborativen**

Wörterbüchern

## Indirekte Beteiligung

- **Explizites Feedback**  
(z.B. Belege)
- **Implizites Feedback**  
durch Wörterbuch-  
verwendung

## Begleitende Beteiligung

- Austausch** zwischen
- Wörterbuchstellern  
und -nutzern
  - Wörterbuchbenutzern  
untereinander

(Abel & Meyer, 2016)

# In diesem Vortrag: Direkte Beteiligung

## Direkte Beteiligung

Beiträge zu

- **offen-kollaborativen,**
  - **kollaborativ-  
institutionellen,**
  - **semi-kollaborativen**
- Wörterbüchern

## Indirekte Beteiligung

- **Explizites Feedback**  
(z.B. Belege)

- **Implizites Feedback**  
durch Wörterbuch-  
verwendung

## Begleitende Beteiligung

**Austausch** zwischen

- Wörterbuchstellern  
und  
Wörterbuchbenutzern  
untereinander



(Abel & Meyer, 2016)

# In diesem Vortrag: Direkte Beteiligung

## Direkte Beteiligung

Artikel neu anlegen,  
verändern, löschen

Beiträge zu

- **offen-kollaborativen,**
- **kollaborativ-  
institutionellen,**
- **semi-kollaborativen**

Wörterbüchern

kein fest vorgegebener Herausgeber,  
z.B. Wiktionary

durch Verlag/Institut betreut,  
z.B. Macmillan Open Dictionary

Moderierte Beiträge  
z.B. LEO

(Abel & Meyer, 2016)

## Direkte Beteiligung

Beiträge zu

- **offen-kollaborativen,**
- **kollaborativ-  
institutionellen,**
- **semi-kollaborativen**

Wörterbüchern

Erforschung des kollaborativen Paradigmas ist nach wie vor hochrelevant:

1. Verdrängen kollaborative Wörterbücher die etablierten Verlagsprodukte wie im Fall von Enzyklopädien?

## Direkte Beteiligung

Beiträge zu

- **offen-kollaborativen,**
- **kollaborativ-  
institutionellen,**
- **semi-kollaborativen**

Wörterbüchern

Erforschung des kollaborativen Paradigmas ist nach wie vor hochrelevant:

1. Verdrängen kollaborative Wörterbücher die etablierten Verlagsprodukte wie im Fall von Enzyklopädien?
2. Lexikographische Innovation im kollaborativen Ansatz?
3. Anwendungen der kollaborativ erarbeiteten Angaben über die Lexikographie hinaus?



## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze

# Ähnliche Artikelstrukturen

Helikoptereltern, die

Wortart: **Pluralwort**  
Gebrauch: **salopp, oft abwertend**  
Häufigkeit: **i** ■ ■ ■ ■ ■

**RECHTSCHREIBUNG**

Worttrennung: He|li|ko|p|ter|el|tern

**BEDEUTUNGSÜBERSICHT**

Eltern, die ihre Kinder aus übertriebener Sorge ständig überwachen

**AUSSPRACHE**

Betonung: *Helikoptereltern*

**HERKUNFT**

wegen der wirkungsvollen Kontrolle eines Geschehens vom Hubschrauber aus

**GRAMMATIK**

Pluraletantum

**BLÄTTERN**

Ihre **Spenden** helfen, Wiktionary zu betreiben. [\[ausblenden\]](#)

## Helikoptereltern

Helikoptereltern (Deutsch) [\[ Bearbeiten \]](#)

**Substantiv** [\[ Bearbeiten \]](#)

**Worttrennung:**

*kein Singular, Plural: He-li-ko-p-ter-el-tern*

**Aussprache:**

IPA: [ heliˈkɔptɐˌʔɛltɐn ]

Hörbeispiele: —

**Bedeutungen:**

[1] **überfürsorgliche** Eltern

**Herkunft:**

**Determinativkompositum** aus den Substantiven *Helikopter* und *Eltern*

**Oberbegriffe:**

[1] **Eltern**

**Beispiele:**

[1] Meine Eltern waren alles andere als *Helikoptereltern*.

[1] „*Helikoptereltern* kreisen ständig über ihren Kindern, um jederzeit eingreifen zu können - wie Hütehunde, die rastlos ihre Herde umrunden und immer in Alarmbereitschaft sind.“<sup>[1]</sup>

**Übersetzungen** [\[ Bearbeiten \]](#)

[\[Ausklappen\]](#)

**Referenzen und weiterführende Informationen:**

[1] Wikipedia-Artikel „*Helikoptereltern*“

[1] Online-Wortschatz-Informationssystem Deutsch - **lexiko** „*Helikoptereltern*“

**Quellen:**

1. ↑ <http://www.br.de/radio/bayern2/gesellschaft/notizbuch/helikoptereltern-kindheit-erziehung-100.html>

# Ähnliche Artikelstrukturen

Helikoptereltern, die

Wortart: **Pluralwort**

Gebrauch: **salopp, oft abwertend**

Häufigkeit: **i** ■ ■ ■ ■ ■

**RECHTSCHREIBUNG**

Worttrennung: He|li|ko|p|ter|el|tern

**BEDEUTUNGSÜBERSICHT**

Eltern, die ihre Kinder aus übertriebener Sorge ständig überwachen

**AUSSPRACHE**

Betonung: *Helikoptereltern*

**HERKUNFT**

wegen der wirkungsvollen Kontrolle eines Geschehens vom Hubschrauber aus

**GRAMMATIK**

Pluraletantum

**BLÄTTERN**

## Helikoptereltern

Ihre **Spenden** helfen, Wiktionary zu betreiben. [\[ausblenden\]](#)

### Helikoptereltern (Deutsch) [\[Bearbeiten\]](#)

#### Substantiv [\[Bearbeiten\]](#)

##### Worttrennung:

kein Singular, Plural: He-li-ko-p-ter-el-tern

##### Aussprache:

IPA: [heliˈkɔptɐˌɛltn]

Hörbeispiele: —

##### Bedeutungen:

[1] **überfürsorgliche** Eltern

##### Herkunft:

**Determinativkompositum** aus den Substantiven *Helikopter* und *Eltern*

##### Oberbegriffe:

[1] **Eltern**

##### Beispiele:

[1] Meine Eltern waren alles andere als *Helikoptereltern*.

[1] „*Helikoptereltern* kreisen ständig über ihren Kindern, um jederzeit eingreifen zu können - wie Hütehunde, die rastlos ihre Herde umrunden und immer in Alarmbereitschaft sind.“<sup>[1]</sup>

##### Übersetzungen [\[Bearbeiten\]](#)

[\[Ausklappen\]](#)

##### Referenzen und weiterführende Informationen:

[1] Wikipedia-Artikel „[Helikoptereltern](#)“

[1] Online-Wortschatz-Informationssystem Deutsch - [lexiko „Helikoptereltern“](#)

##### Quellen:

1. ↑ <http://www.br.de/radio/bayern2/gesellschaft/notizbuch/helikoptereltern-kindheit-erziehung-100.html>

# Interessant: Übersetzungen, Gebrauch, Quellen



Ihre [Spenden](#) helfen, Wiktionary zu betreiben. [\[ausblenden\]](#)

Helikoptereltern, die

Wortart: **Pluralwort**  
Gebrauch: **salopp, oft abwertend**  
Häufigkeit:

**RECHTSCHREIBUNG**

Worttrennung: He|li|ko|p|ter|el|tern

**BEDEUTUNGSÜBERSICHT**

Eltern, die ihre Kinder aus übertriebener Sorge ständig überwachen

**AUSSPRACHE**

Betonung: *Helikoptereltern*

**HERKUNFT**

wegen der wirkungsvollen Kontrolle eines Geschehens vom Hubschrauber aus

**GRAMMATIK**

Pluraletantum

**BLÄTTERN**

2

1

3

## Helikoptereltern

Helikoptereltern (Deutsch) [\[ Bearbeiten \]](#)

**Substantiv** [\[ Bearbeiten \]](#)

**Worttrennung:**  
*kein Singular, Plural: He-li-ko-p-ter-el-tern*

**Aussprache:**  
IPA: [heliˈkɔptɐˌɛltn]  
Hörbeispiele: —

**Bedeutungen:**  
[1] **überfürsorgliche** Eltern

**Herkunft:**  
Determinativkompositum aus den Substantiven *Helikopter* und *Eltern*

**Oberbegriffe:**  
[1] *Eltern*

**Beispiele:**  
[1] Meine Eltern waren alles andere als *Helikoptereltern*.  
[1] „*Helikoptereltern* kreisen ständig über ihren Kindern, um jederzeit eingreifen zu können - wie Hütehunde, die rastlos ihre Herde umrunden und immer in Alarmbereitschaft sind.“<sup>[1]</sup>

	Singular	Plural
<b>Nominativ</b>	—	die Helikoptereltern
<b>Genitiv</b>	—	der Helikoptereltern
<b>Dativ</b>	—	den Helikoptereltern
<b>Akkusativ</b>	—	die Helikoptereltern

**Übersetzungen** [\[ Bearbeiten \]](#)

[\[Ausklappen\]](#)

**Referenzen und weiterführende Informationen:**

- [1] Wikipedia-Artikel „[Helikoptereltern](#)“
- [1] Online-Wortschatz-Informationssystem Deutsch - [lexiko „Helikoptereltern“](#)

**Quellen:**

1. ↑ <http://www.br.de/radio/bayern2/gesellschaft/notizbuch/helikoptereltern-kindheit-erziehung-100.html>





DE  
|  
EN

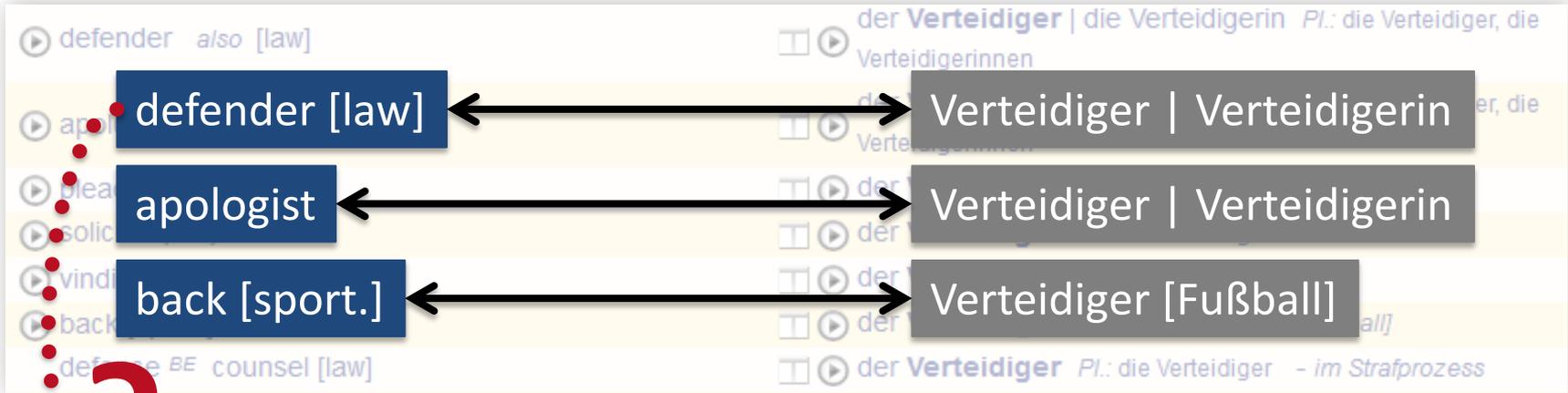
▶ defender <i>also</i> [law]	☐ ▶ der <b>Verteidiger</b>   die Verteidigerin <i>Pl.:</i> die Verteidiger, die Verteidigerinnen
▶ apologist	☐ ▶ der <b>Verteidiger</b>   die Verteidigerin <i>Pl.:</i> die Verteidiger, die Verteidigerinnen
▶ pleader	☐ ▶ der <b>Verteidiger</b> <i>Pl.:</i> die Verteidiger
▶ solicitor ( <i>Brit.</i> )	☐ ▶ der <b>Verteidiger</b> <i>Pl.:</i> die Verteidiger
▶ vindicator	☐ ▶ der <b>Verteidiger</b> <i>Pl.:</i> die Verteidiger
▶ back [sport.]	☐ ▶ der <b>Verteidiger</b> <i>Pl.:</i> die Verteidiger [ <i>Fußball</i> ]
defence <sup>BE</sup> counsel [law]	☐ ▶ der <b>Verteidiger</b> <i>Pl.:</i> die Verteidiger - <i>im Strafprozess</i>

DE  
|  
ES

▶ el defensor   la defensora	☐ ▶ der <b>Verteidiger</b>   die Verteidigerin <i>Pl.:</i> die Verteidiger, die Verteidigerinnen
▶ el apologista   la apologista	☐ ▶ der <b>Verteidiger</b>   die Verteidigerin <i>Pl.:</i> die Verteidiger, die Verteidigerinnen
▶ abogado defensor [jur.]	☐ ▶ der <b>Verteidiger</b>   die Verteidigerin <i>Pl.:</i> die Verteidiger, die Verteidigerinnen
▶ el líbero [dep.]	☐ ▶ der Libero <i>Pl.:</i> die Liberos <i>italienisch</i> [ <i>Fußball</i> ] - <i>freier Verteidiger</i>

# Übersetzungstabellen für Sprachpaare

DE  
|  
EN



defender also [law]      der **Verteidiger** | die Verteidigerin Pl.: die Verteidiger, die Verteidigerinnen

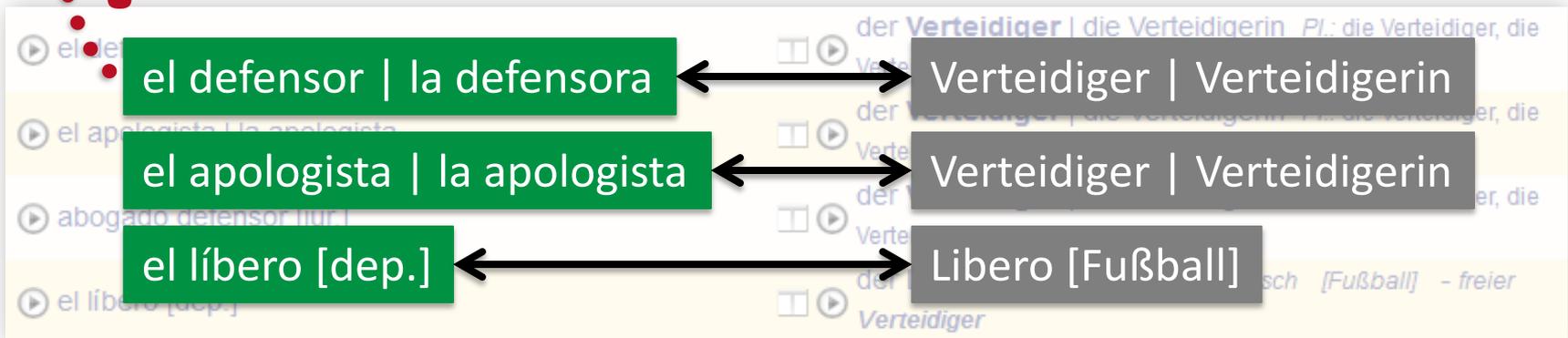
apologist      der **Verteidiger** | die Verteidigerin

back [sport.]      der **Verteidiger** [Fußball]

defence BE counsel [law]      der **Verteidiger** Pl.: die Verteidiger - im Strafprozess



DE  
|  
ES



el defensor | la defensora      der **Verteidiger** | die Verteidigerin

el apologista | la apologista      der **Verteidiger** | die Verteidigerin

el líbero [dep.]      **Libero** [Fußball]

# Refugee Phrasebook



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

German	English	Syrian/Arabic alphabet	Spanish	Bangla	Urdu
Richter	judge	القاضي	El juez / La jueza	বিচারক	اب کو جج کی ضرورت ہے؟
Rückübernahmeabkommen	re-admission agreement (between two countries)	اتفاقية إعادة القبول (بين دولتين)	Acuerdo de readmisión	পুনরায় দেশে ঢোকার চুক্তি (দুই দেশের মধ্যে)	دوباره داخلے معاہدہ
subsidiärer Schutz nach EU Recht	subsidiary protection under EU law	الحماية الثانوية في قانون الاتحاد الأوروبي (أي الحماية التي تطل رعايا دول أخرى)	protección subsidiaria bajo legislación europea	ইউরোপিয়ান ইউনিয়নের সুরক্ষা আইন	
Übersetzung	Translation	الترجمة	La traducción	অনুবাদ	ترجمہ
Verhaftung	arrest/detention	التوقيف / الاعتقال	La detención / el arresto	গ্রেফতার / আটক	گرفتار
Visum	visa	تأشيرة دخول	Visa / visado	ভিসা	ویزا
Visablehnung	refusal of visa application	رفض تأشيرة الدخول	Denegación de visado	ভিসার আবেদন নাকচ	ویزا کی درخواست سے انکار
Wiedererwägung	reconsideration	إعادة نظر	La reconsideración	পুনরায় আবেদনপত্র যাচাই।	
Pflichtverteidiger	assigned counsel/duty counsel	محامي بالمجان	Abogado de oficio	নির্ধারিত/ডিউটিতে থাকা পরামর্শক	
Verteidiger	counsel for defence	محامي الدفاع	Abogado defensor	বিবাদীর উকিল	
Freispruch	acquittal	الحكم بالبراءة	Veredicto absolutorio	দোষ থেকে মুক্তি	بری

# Mehrsprachige Übersetzungstabelle

German	English	Syrian/Arabic alphabet	Spanish	Bangla	Urdu
Richter	judge	القاضي	El juez / La jueza	বিচারক	اب کو جج کی ضرورت ہے؟
Rückübernahmeabkommen	re-admission agreement (between two countries)	اتفاقية إعادة القبول (بين دولتين)	Acuerdo de readmisión	পুনরায় দেশে ঢোকার চুক্তি (দুই দেশের মধ্যে)	دوباره داخلے معاہدہ
subsidiärer Schutz nach EU Recht	subsidiary protection under EU law	الحماية الثانوية في قانون الاتحاد الأوروبي (أي الحماية التي تطال رعايا دول أخرى)	protección subsidiaria bajo legislación europea	ইউরোপিয়ান ইউনিয়নের সুরক্ষা আইন	
Übersetzung					
	Richter	judge	القاضي	El juez / La jueza	
Verhaftung	arrest/detention	التوقيف / الاعتقال	La detención / el arresto	গ্রেফতার / আটক	گرفتار
Visum	visa	تأشيرة دخول	تأشيرة دخول	ভিসা	ویزا
Visablehnung	refusal of visa application	رفض تأشيرة الدخول	Denegación de visado	ভিসার আবেদন নাকচ	ویزا کی درخواست سے انکار
Wiedererwägung	reconsideration	إعادة نظر	La reconsideración	পুনরায় আবেদনপত্র যাচাই	
Pflichtverteidiger	counsel for defence	محامي الدفاع	محامي الدفاع	নির্ধারিত বিবাদীর রক্ষাকর্তা	
Verteidiger	counsel for defence	محامي الدفاع	Abogado defensor	বিবাদীর উকিল	
Freispruch	acquittal	الحكم بالبراءة	Veredicto absolutorio	দোষ থেকে মুক্তি	بری

# Feste Quellsprache (Wiktionary)

Verteidiger (Deutsch) [Bearbeiten]

## Bedeutungen:

- [1] *Sport, Fußball*: defensiv eingestellter Spieler, dessen Aufgabe es ist, das gegnerische Spiel zu stören, um anschließend das eigene Spiel wieder aufnehmen zu können
- [2] *Recht*: Rechtsanwalt, der vor Gericht die Interessen des Angeklagten vertritt
- [3] *Militär, auch übertragen in Spielen*: Gegenpart zum Angreifer, in der Regel in der eigenen Abwehrlinie
- [4] *Abwehr*: was (zum Beispiel eine Idee) gegen etwas (zum Beispiel einen Angriff) wehrt

Übersetzung

- Bulgarisch: [1] *бранител* → *bg*
- Englisch: [1] *back* → *en*; [2] *defence* → *en*
- Französisch: [1] *défenseur* → *fr* *m*; [2] *avocat de la défense* → *fr* *m*
- Isländisch: [1] *verjandi* → *is*
- Italienisch: [1] *difensore* → *it* *m*
- Katalanisch: [1] *defensa* → *ca*
- Litauisch: [1] *gynėjas* → *lt*
- Niederländisch: [1, 2] *verdediger* → *nl* *m, f*
- Okzitanisch: [1] *arrièr* → *oc*
- Polnisch: [1, 2] *obrońca* → *pl* *m*, [1] *defensor* → *pl* *m*, [2] *adwokat* → *pl* *m*
- Russisch: [1, 2] *защитник* (*zašitnik*) → *ru* *m*
- Schwedisch: [1] *back* → *sv* *u*; [2] *försvarsadvokat* → *sv* *u*
- Spanisch: [1] *defensa* → *es* *m, f*, [2] *abogado* → *es* *m*

Verteidiger<sub>1</sub>

Verteidiger<sub>2</sub>

Verteidiger<sub>3</sub>

Verteidiger<sub>4</sub>

back

defensa

защитник

defence

abogado

Verteidigung

Wehr



# Mehrsprachige Synsets (OmegaWiki)

## Verteidiger

Sprache: Deutsch

► **Verteidiger** : One who supports something.

▼ **Verteidiger** : Someone who defends with conviction one thesis, ideal or plan.

▼ Definitionen

Verteidiger <sub>1</sub>	Befürworter <sub>1</sub>	Vertreter <sub>1</sub>	partidaria <sub>1</sub>
--------------------------	--------------------------	------------------------	-------------------------

▼ Sprachliche Texten

Englisch	advocate <sub>1</sub>	proponent <sub>1</sub>	partidario <sub>1</sub>	proponente <sub>1</sub>
----------	-----------------------	------------------------	-------------------------	-------------------------

Italienisch: Persona che difende con convinzione una tesi, un ideale, un progetto.

Niederländisch: Iemand die met overtuiging een stelling, ideaal of plan verdedigd.

Spanisch

Verteidiger <sub>2</sub>	Verfechter <sub>2</sub>	defensor <sub>1</sub>	defensora <sub>1</sub>
--------------------------	-------------------------	-----------------------	------------------------

▼ Synonyme und Übersetzungen

Ausdruck	defender <sub>1</sub>	supporter <sub>1</sub>	защитник <sub>1</sub>
----------	-----------------------	------------------------	-----------------------

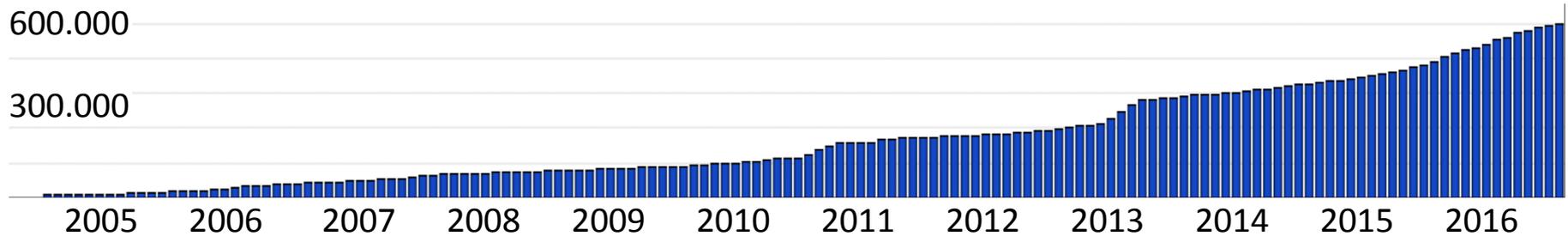
▼ Sprachliche Anmerkungen

Arabisch	مدعم	Anzeigen ▼
Bokmål	forsvarer	Anzeigen ▼
Bulgarisch	защитник	Anzeigen ▼
Dänisch	forsvarer	Anzeigen ▼
Deutsch	Verfechter	Anzeigen ▼
Englisch	defender	Anzeigen ▼

## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze

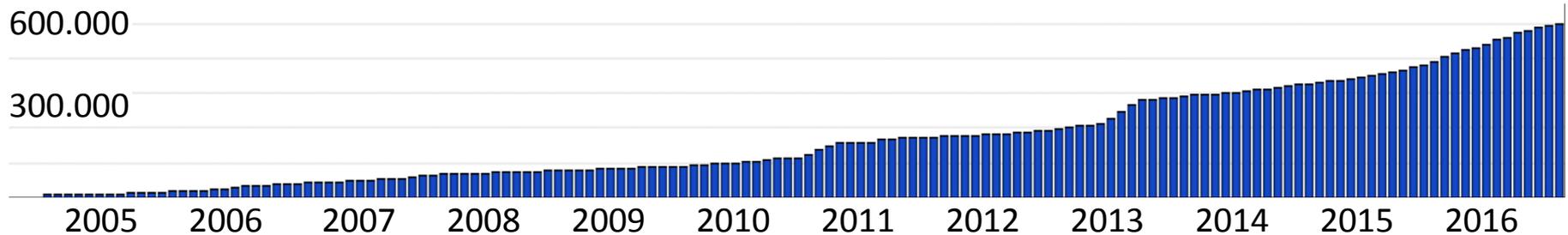
# Entwicklung der Artikelzahlen

## Anzahl Wörterbuchartikel

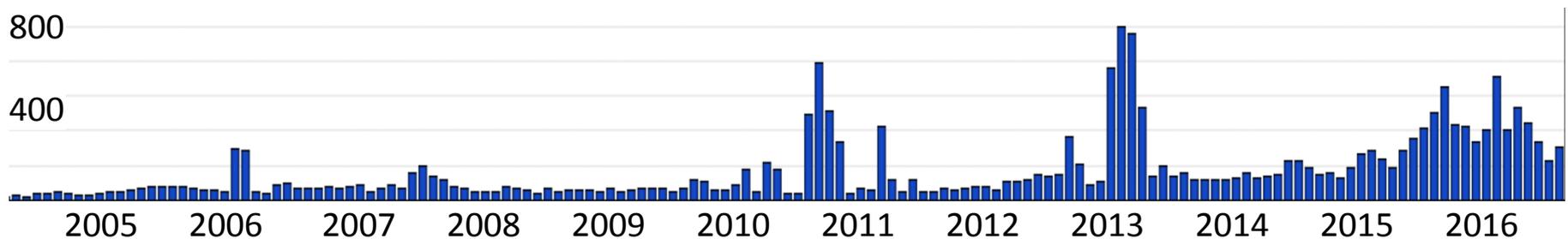


# Entwicklung der Artikelzahlen

## Anzahl Wörterbuchartikel



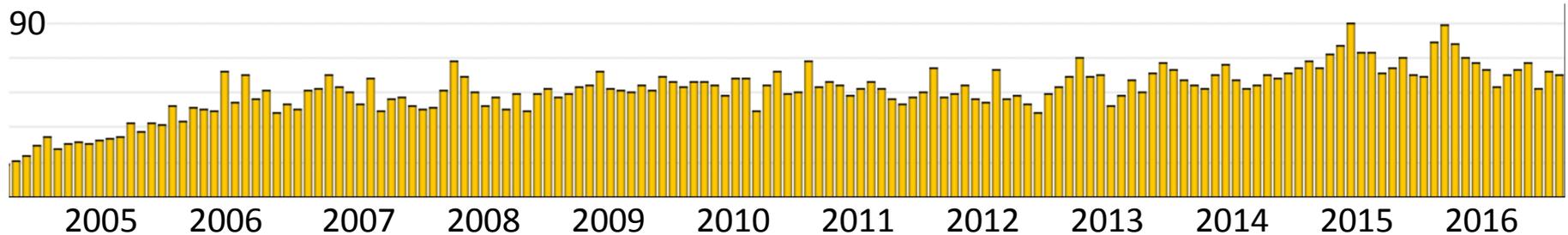
## Neue Artikel pro Tag



<https://stats.wikimedia.org/wiktionary/DE/ChartsWikipediaDE.htm> (public domain, Stand: 07.03.2017)

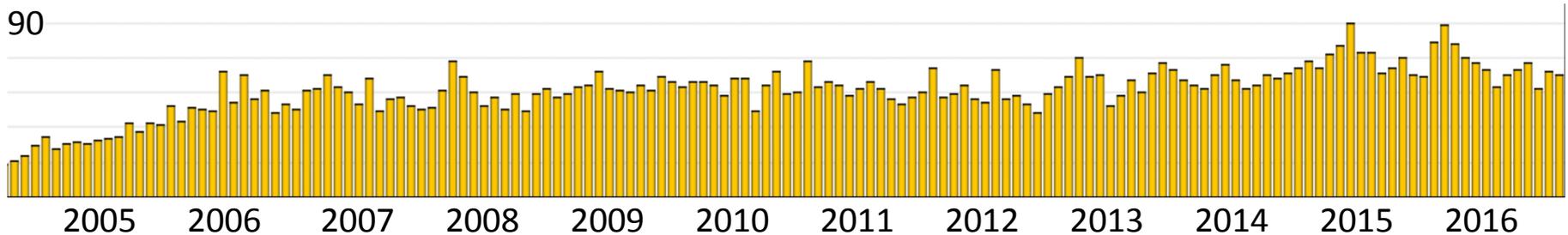
# Entwicklung der Beitragenden

## Aktive Autor\_innen (> 5 Änderungen; von $\approx 1.500$ registrierten)



# Entwicklung der Beitragenden

## Aktive Autor\_innen (> 5 Änderungen; von ≈ 1.500 registrierten)



### Einige Nutzerrollen:

- Fremdsprachiges
- Fachthemen
- Artikelformatierung
- Vandalismusbekämpfung

„Hauptautor\_in“: ca.  $\frac{1}{3}$  der Überarbeitungen eines Artikels

**Bots:** Beigetragene Programme für automatisierbare Aufgaben (z.B. für konsistentes Layout)

# Kollaborativ Lexikographischer Prozess

*Vorbereitungsphase*



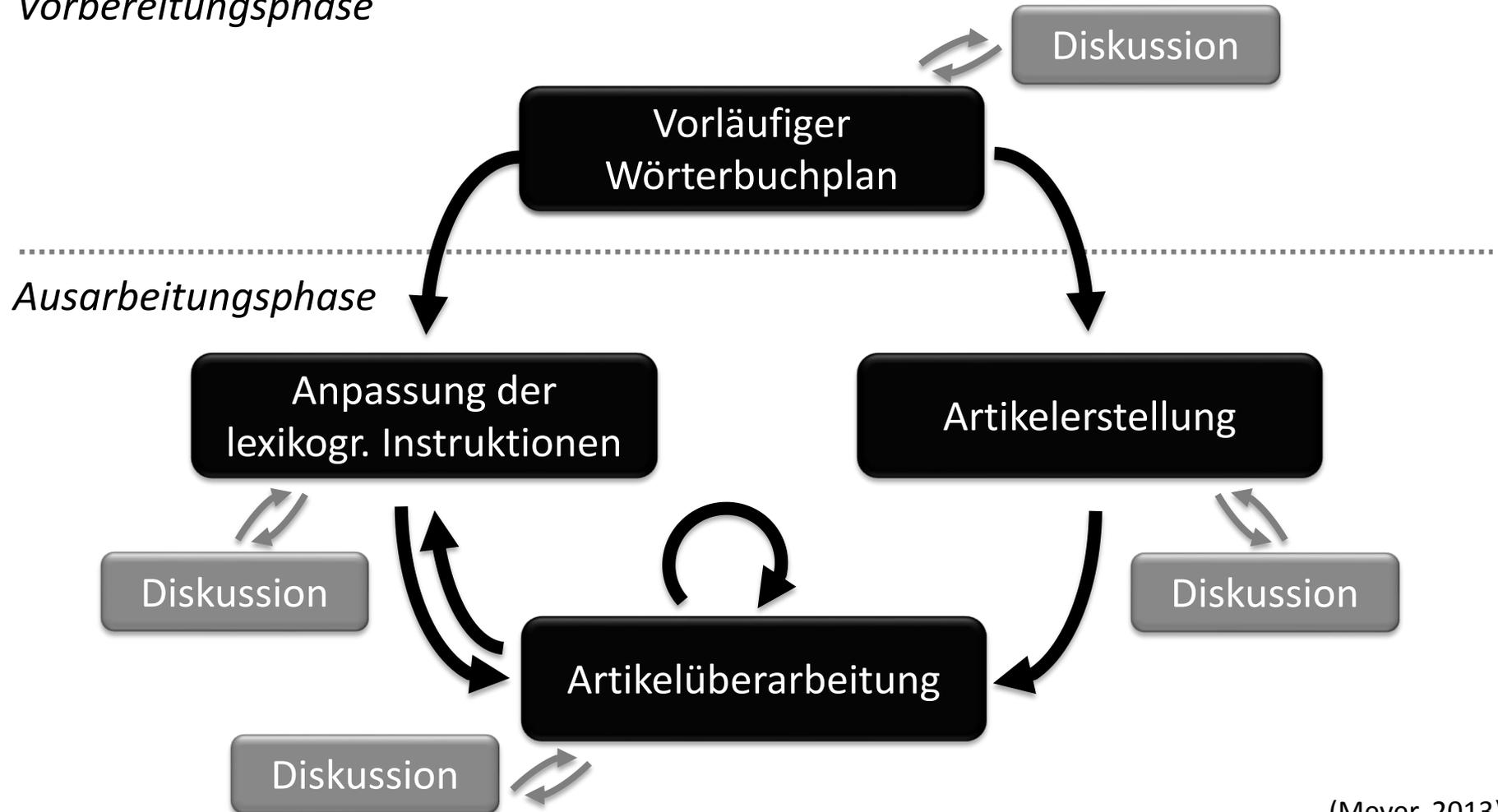
*Ausarbeitungsphase*



(Meyer, 2013)

# Kollaborativ Lexikographischer Prozess

*Vorbereitungsphase*



(Meyer, 2013)



## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze



# JWCTL: Java-based Wiktionary Library

- Strukturierter Zugriff auf Angaben im englischen, deutschen und russischen Wiktionary
- Wortarten
- Bedeutungsparaphrasen
- Semantische Relationen
- Übersetzungen

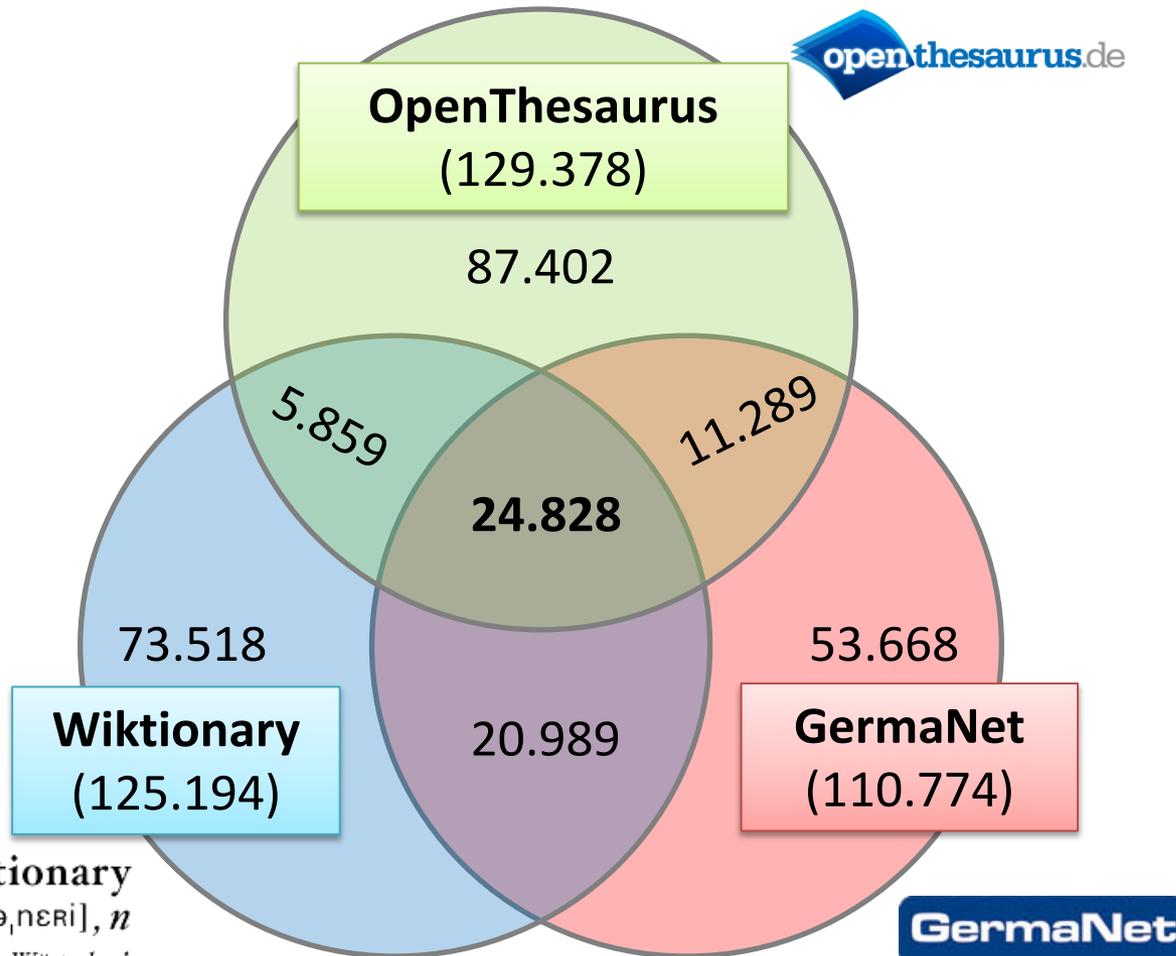


## Quelloffene Software:

<https://www.ukp.tu-darmstadt.de/software/jwctl/>

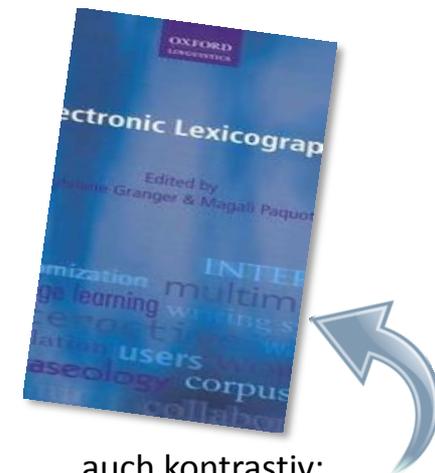
<https://dkpro.github.io/dkpro-jwctl/>

# Quantitativer Wörterbuchvergleich



**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*

Stand: 02/2017 (Wiktionary), 04/2014 (GermaNet), 03/2017 (OpenThesaurus); nur vergleichbare Einträge beachtet.



auch kontrastiv:  
Meyer & Gurevych (2012)

# Unterschiedliche Teilwortschätze

**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*



**Verben:** anhusten,  
banalisieren, fertig  
bekommen, lasten,  
neufassen, rümpfen,...

**Zahlwort:** 100-fach,  
achtspurig, elfbändig,  
neunmonatig,...

**Berufsbezeichnungen:**  
Altenpflegehelferin,  
Erzkanzler, Mälzerin,...

# Unterschiedliche Teilwortschätze

**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*



## Neues/Aktuelles:

Flexirente, Flüchtlings\*  
(97x), Helikoptereltern,  
Inklusionsklasse,  
Schmähgedicht,...

**Verben:** anhusten,  
banalisieren, fertig  
bekommen, lasten,  
neufassen, rümpfen,...

**Zahlwort:** 100-fach,  
achtspurig, elfbändig,  
neunmonatig,...

**Berufsbezeichnungen:**  
Altenpflegehelferin,  
Erzkanzler, Mälzerin,...

# Unterschiedliche Teilwortschätze

**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*



## **Neues/Aktuelles:**

Flexirente, Flüchtlings\*  
(97x), Helikopterneltern,  
Inklusionsklasse,  
Schmähgedicht,...

**Fachliches:** Adjektiv-  
phrase, Gichtgas,  
Qualitätsopfer,...

**Fachliches:** Editions-  
philologie, Grafik-  
treiber, Pandektistik,...

**Verben:** anhusten,  
banalisieren, fertig  
bekommen, lasten,  
neufassen, rümpfen,...

**Zahlwort:** 100-fach,  
achtspurig, elfbändig,  
neunmonatig,...

**Berufsbezeichnungen:**  
Altenpflegehelferin,  
Erzkanzler, Mälzerin,...

# Unterschiedliche Teilwortschätze

**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*



## **Neues/Aktuelles:**

Flexirente, Flüchtlings\*  
(97x), Helikoptereltern,  
Inklusionsklasse,  
Schmähgedicht,...

**Fachliches:** Adjektiv-  
phrase, Gichtgas,  
Qualitätsopfer,...

**Umgangssprachliches:**  
Gurtmuffel, Hupfdohle

**Fachliches:** Editions-  
philologie, Grafik-  
treiber, Pandektistik,...

**Umgangssprachliches:**  
Asi-Toaster, Ecken-  
pinkler, Kneipenkaiser,  
Lauschlappen, rappel-  
voll, schäbbig [sic!],...

**Verben:** anhusten,  
banalisieren, fertig  
bekommen, lasten,  
neufassen, rümpfen,...

**Zahlwort:** 100-fach,  
achtspurig, elfbändig,  
neunmonatig,...

**Berufsbezeichnungen:**  
Altenpflegehelferin,  
Erzkanzler, Mälzerin,...

# Unterschiedliche Teilwortschätze

**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*



## **Neues/Aktuelles:**

Flexirente, Flüchtlings\*  
(97x), Helikoptereltern,  
Inklusionsklasse,  
Schmähgedicht,...

**Fachliches:** Adjektiv-  
phrase, Gichtgas,  
Qualitätsopfer,...

**Umgangssprachliches:**  
Gurtmuffel, Hupfdohle

**Fachliches:** Editions-  
philologie, Grafik-  
treiber, Pandektistik,...

**Umgangssprachliches:**  
Asi-Toaster, Ecken-  
pinkler, Kneipenkaiser,  
Lauschlappen, rappel-  
voll, schäbbig [sic!],...

**Phrasen:** *(etwas) den  
Vorzug geben,...*

**Verben:** anhusten,  
banalisieren, fertig  
bekommen, lasten,  
neufassen, rümpfen,...

**Zahlwort:** 100-fach,  
achtspurig, elfbändig,  
neunmonatig,...

**Berufsbezeichnungen:**  
Altenpflegehelferin,  
Erzkanzler, Mälzerin,...

# Das schwierige Thema Qualität

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge

„Wörterbuchqualität“

Dieser Eintrag existiert noch nicht.

# Das schwierige Thema Qualität

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge
- Plagiarismus und automatische Übernahme

„Wörterbuchqualität“

Dieser Eintrag existiert noch nicht.

# Das schwierige Thema Qualität

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge
- Plagiarismus und automatische Übernahme

„Wörterbuchqualität“

Dieser Eintrag existiert noch nicht.

## Aber: überhaupt vergleichbar?

- „bewegliches Ziel“: evolutionäre Entwicklung
- Maßnahmen zu Qualitätssicherung
- Sammeln vs. Erarbeiten
- Introspektion vs. Korpusevidenz

(Fuertes-Olivera, 2009; Hanks, 2012; Rundell, 2012; Meyer, 2013; EURALEX-Workshop, 2014; Meyer & Gurevych, 2014)

# Unterschiedliche Quellen

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge
- Plagiarismus und automatische Übernahme

## Aber: überhaupt vergleichbar?

- „bewegliches Ziel“: evolutionäre Entwicklung
- Maßnahmen zu Qualitätssicherung
- Sammeln vs. Erarbeiten
- Introspektion vs. Korpusvidenz

## Korpusvidenz:

- Kein Projekt mit eigenem Kernkorpus
- Keine Instruktionen zur systematischen Analyse

(Fuertes-Olivera, 2009; Hanks, 2012; Rundell, 2012; Meyer, 2013; EURALEX-Workshop, 2014; Meyer & Gurevych, 2014)

# Unterschiedliche Quellen

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge
- Plagiarismus und automatische Übernahmen

## Aber: überhaupt vergleichbar?

- „bewegliches Ziel“: evolutionäre Entwicklung
- Maßnahmen zu Qualitätssicherung
- Sammeln vs. Erarbeiten
- Introspektion vs. Korpusevidenz

## Korpusevidenz:

- Kein Projekt mit eigenem Kernkorpus
- Keine Instruktionen zur systematischen Analyse

## Quellen:

- oft Wörterbücher
- zuweilen Webquellen und Fachliteratur

(Fuertes-Olivera, 2009; Hanks, 2012; Rundell, 2012; Meyer, 2013; EURALEX-Workshop, 2014; Meyer & Gurevych, 2014)

# Unterschiedliche Quellen

## Fragwürdige Qualität in Wiktionary

- Lücken
- Altertümliche Beschreibungen
- Fehlerhafte Einträge
- Plagiarismus und automatische Übernahmen

## Aber: überhaupt vergleichbar?

- „bewegliches Ziel“: evolutionäre Entwicklung
- Maßnahmen zu Qualitätssicherung
- Sammeln vs. Erarbeiten
- Introspektion vs. Korpusevidenz

## Korpusevidenz:

- Kein Projekt mit eigenem Kernkorpus
- Keine Instruktionen zur systematischen Analyse

## Quellen:

- oft Wörterbücher
- zuweilen Webquellen und Fachliteratur
- Überlegungen auf Diskussionsseiten

(Fuertes-Olivera, 2009; Hanks, 2012; Rundell, 2012; Meyer, 2013; EURALEX-Workshop, 2014; Meyer & Gurevych, 2014)



## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze

1. Verdrängen kollaborative Wörterbücher die etablierten Verlagsprodukte wie im Fall von Enzyklopädien?
2. Lexikographische Innovation im kollaborativen Ansatz?
3. Nutzen der kollaborativ erarbeiteten Angaben über die Lexikographie hinaus?

## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze

1. Verdrängen kollaborative Wörterbücher die etablierten Verlagsprodukte wie im Fall von Enzyklopädien?
2. Lexikographische Innovation im kollaborativen Ansatz?
3. Nutzen der kollaborativ erarbeiteten Angaben über die Lexikographie hinaus?



## Strukturen, Dynamik und Zusammensetzung gemeinschaftlich erarbeiteter Wortschätze

1. Verdrängen kollaborative Wörterbücher die etablierten Verlagsprodukte wie im Fall von Enzyklopädien?
2. Lexikographische Innovation im kollaborativen Ansatz?
3. Nutzen der kollaborativ erarbeiteten Angaben über die Lexikographie hinaus?

# Anwendung: Interkulturelles Marketing

---

(Meyer et al., 2016)

# Anwendung: Interkulturelles Marketing



**Produkt-/Markenname**  
z.B. Silver Mist



**Problematische Bedeutungen  
in anderen Sprachen**  
z.B. Mist (Deutsch)

# Anwendung: Interkulturelles Marketing



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



**Produkt-/Markenname**  
z.B. Silver Mist

## Übersetzungen

**Wiktionary** in zahlreiche Sprachen  
[ˈvɪkʃəˌnɛʀi], *n*  
*Das freie Wörterbuch*

## Sprachvarietäten

insbes. Umgangssprache,  
Jargon, Neuschöpfungen



**Problematische Bedeutungen  
in anderen Sprachen**  
z.B. Mist (Deutsch)

# Evaluation im Kurzüberblick

## Experiment 1: Datensatz aus Ricks (2006)

	MARKED	SOUNDEX	LOOKUP	COMBINE
Detected blunders:	18 / 44	18 / 44	28 / 44	34 / 44
Relevant clues:	105 / 151	85 / 247	341 / 1202	229 / 517
Precision $P$ :	<b>.70</b>	.34	.28	.44
Recall $R$ :	.41	.41	<b>.64</b>	<b>.77</b>
$F_1$ score:	.52	.37	.39	<b>.56</b>
$F_2$ score:	.45	.39	.51	<b>.67</b>

**Experiment 2: ca. 1.000 internationale Marken aus BrandPitt corpus**  
(Özbal et al., 2012): 154 bzw. 192 relevante Hinweise (MARKED)

# Vielen Dank für Ihre Aufmerksamkeit!

- A. Abel, Ch.M. Meyer (2016): Nutzerbeteiligung.  
In: *Internetlexikografie. Ein Kompendium*,  
S. 249–290. De Gruyter.
- P.A. Fuertes-Olivera (2009): The Function Theory  
of Lexicography and Electronic Dictionaries:  
Wiktionary as a Prototype of Collective Free  
Multiple-Language Internet Dictionary. In:  
*Lexicography at a Crossroads: Dictionaries and  
Encyclopedias Today, Lexicographical Tools  
Tomorrow*, S. 99–134. Peter Lang.
- P. Hanks (2012): Word Meaning and Word Use:  
Corpus evidence and electronic lexicography.  
In: *Electronic Lexicography*, S. 57–82. OUP.
- Ch.M. Meyer (2013): *Wiktionary: The Meta-  
lexicographic and the Natural Language  
Processing Perspective*. tprints 3654.
- Ch.M. Meyer, J. Eckle-Kohler, I. Gurevych (2016):  
Semi-automatic Detection of Cross-lingual  
Marketing Blunders based on Pragmatic Label  
Propagation in Wiktionary. In: *Proceedings of  
COLING*, S. 2071–2081.
- Ch.M. Meyer, I. Gurevych (2012): Wiktionary: A  
new rival for expert-built lexicons? Exploring the  
possibilities of collaborative lexicography.  
In: *Electronic Lexicography*, S. 259–291. OUP.
- Ch.M. Meyer, I. Gurevych (2014): Methoden  
bei kollaborativen Wörterbüchern,  
*Lexicographica* 30(1): 187–212.
- M. Rundell: ‘It works in practice but will it work in  
theory?’ The uneasy relationship between  
lexicography and matters theoretical. In:  
*Proceedings of EURALEX*, S. 47–92.
- K. Thier (2014): Das Oxford English Dictionary und  
seine Nutzer. In: *Ihr Beitrag bitte! – Der Nutzer-  
beitrag im Wörterbuchprozess. 3. Arbeitsbericht  
des wissenschaftlichen Netzwerks „Internet-  
lexikografie“*. IDS Mannheim.
- Was ist ein gutes (Internet-)Wörterbuch? – Alte und  
neue Fragen zur Qualität lexikographischer  
Produkte im digitalen Zeitalter* (2014). EURALEX  
Pre-conference Workshop, Bozen/Bolzano,  
11.07.2014. <http://internetlexikografie.de>.

## Kontakt / Contact

### Dr. Christian M. Meyer

Technische Universität Darmstadt  
Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16–25293

 +49 (0)6151 16–25295

 meyer (at) ukp.informatik.tu-darmstadt.de

### Rechtliche Hinweise

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

### Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.