

Qualität und Qualitätssicherung in kollaborativen Wörterbüchern



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Christian M. Meyer

Linguistische Forschung an der FAU: „Online-Wörterbücher und Grammatik“,
Interdisziplinäres Zentrum für Lexikografie, Valenz- und Kollokationsforschung,
Friedrich-Alexander-Universität Erlangen-Nürnberg, 25. November 2014.

de|di|zie|ren <sw. V.; hat> [lat. dedicare, eigtl. = überliefernd kundgeben] (bildungsspr.): *jmdm. zueignen, widmen, für ihn bestimmen*: der Autor dedizierte das Buch seinem Lehrer.

de|zi|die|ren <sw. V.; hat> [lat. decidere, eigtl. = abschneiden, aus: de- = von – weg u. caedere, ↑ Zäsur] (bildungsspr.): *entscheiden; bestimmen*.

Wir brauchen Wörterbücher!

- | Form
- | formal
- | formell
- | förmlich
- | Forscher
- | Forscherin
- | Forschung
- | fortgesetzt
- | Fortschritt
- | **Foto**
- | Frage
- | fragen
- | Fraktion
- | Fraktionschef
- | Fraktionsvorsitzende
- | Fraktionsvorsitzender

Fotobombe



Wir brauchen aktualisierte Wörterbücher!

Brücke (Deutsch) [Bearbeiten]

Substantiv, f [Bearbeiten]

Bedeutungen:

- [1] ein künstlicher Weg zur Überquerung
- [2] *Zahnmedizin*: modellierte Zahnbauform
- [3] ein länglicher Teppich
- [4] *Anatomie*: Teil des Wirbels
- [5] *übertragen*: die Verbrückung
- [6] eine Turnübung
- [7] *Schifffahrt*, kurz : Kommandobrücke

?



Wir brauchen hilfreiche Wörterbücher!

Wie lange dauert die Ausarbeitung?

- Wörterbuchprojekte sind enorm aufwendig und meist langfristig angelegt
- Beispiele für lexikographische Großprojekte:
 - OED (1857–1928)
 - DWB 1854–1961)
 - DWB Neubearbeitung (1965–2016?)
 - WBÖ (1963–2020?)
 - DFD (2012–2034?)

Wie steht es um die Finanzierung?

- Medienkrise: angespannte Situation bei Wörterbuchverlagen
- Diskussion zu Akademievorhaben

Das Internet als Interaktions- und Mitmachplattform

Smartphone

Phablet

Tablet

Notebook

Desktop-PC



Soziales Netzwerk

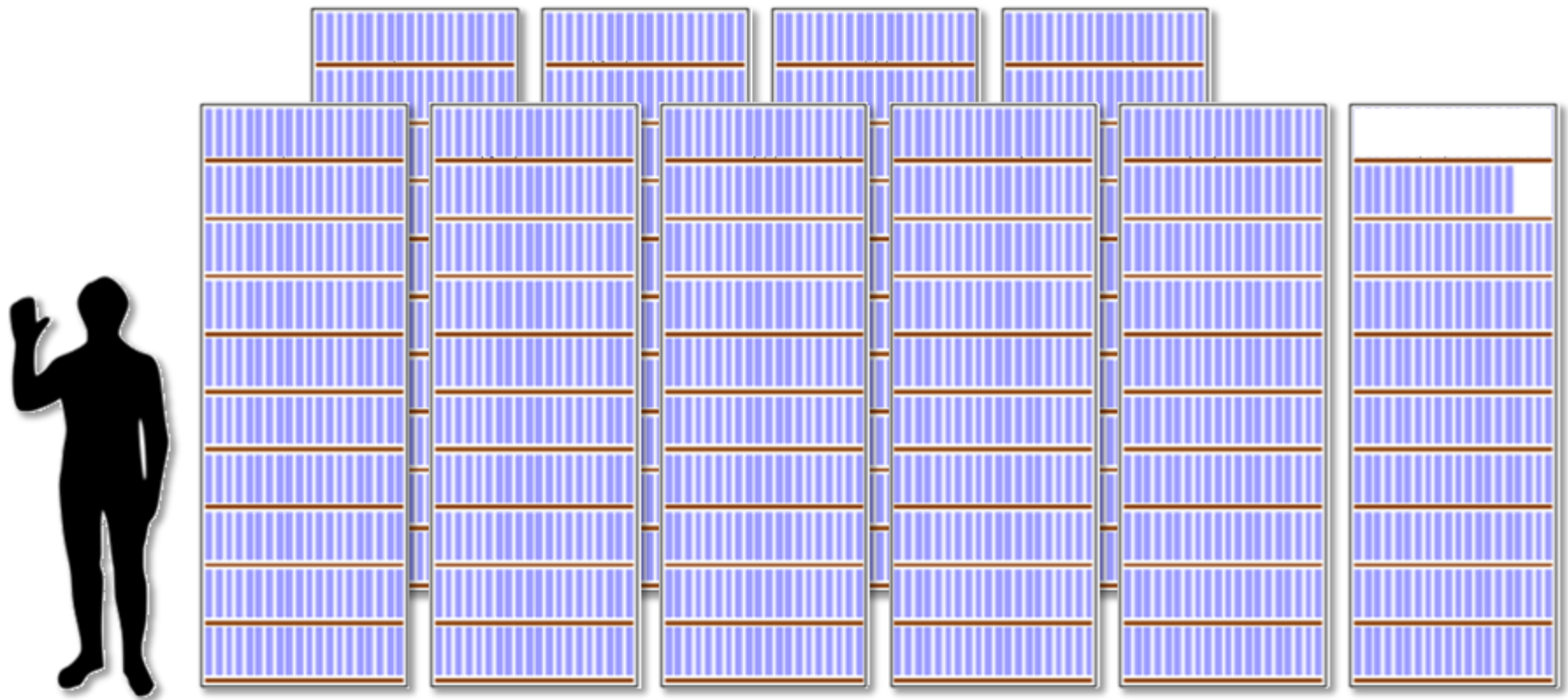
Forum

Wiki

Weblog

Soziales Taggen

Wikipedia: ≈ 1.976 Bände, 10 Bücherregale (Stand März 2014)



http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

Wiktionary – Das Mitmachwörterbuch



TECHNISCHE
UNIVERSITÄT
DARMSTADT

encyclopedia

Wiktionary

['wɪkʃənɹɪ] *n.*,
a wiki-based Open
Content dictionary

Wileo [ˈwɪl kəɹɪ]

Main Page

Community portal

Preferences

Requested entries

Recent changes

Random entry
(by language)

Help

Donations

Contact us

▼ Toolbox

What links here

Related changes

Upload file

Special pages

Printable version

Permanent link

Page information

Cite this page

Add definition

▼ In other projects

Wikipedia

▼ Visibility

Entry

Discussion

Citations

Read

Edit

History

Search

Q

boat

English

[edit]

Etymology

[edit]

From Middle English *boot*, *bot*, *boet*, *boyt* ("boat"), from Old English *bāt* ("boat"), from Proto-Germanic **baltaz*, **baltan* ("boat, small ship"), from Proto-Indo-European **bheid-* ("to break, split"). Cognate with Old Norse *beiti* ("boat").

Old Norse *bātr* (whence Icelandic *bátur*, Norwegian *bått*), Dutch *boot*, German *Boot*, and French *bateau* are all ultimately borrowings from the Old English word.

Pronunciation

[edit]

- (RP) enPR: bōt, IPA: /bəʊt/, X-SAMPA: /bəʊt/
- Rhymes: -əʊt
- (GenAm) enPR: bōt, IPA: /boʊt/, X-SAMPA: /boʊt/
- Rhymes: -oʊt
- Audio (US) 

Noun

[edit]

boat (plural **boats**)

1. A craft used for transportation of goods, fishing, racing, recreational cruising, or military use on or in the water, propelled by oars or outboard motor or inboard motor or by wind.
2. (*poker slang*) A full house.
3. (*chemistry*) One of two possible conformations of cyclohexane rings (the other being chair), shaped roughly like a boat.

Usage notes

[edit]

There's no explicit limit, but the word **boat** usually refers to a relatively small watercraft that is generally smaller than a "ship" and

 Wikipedia has an article on:
Boat



A boat kept on land

25.11.2014 | Fachbereich Informatik | UKP Lab – Prof. Dr. Iryna Gurevych | Christian M. Meyer | 8

 UKP

Neues Paradigma der Wörterbucharstellung: Kollaborative Lexikographie

- Wikipedia verdrängt zunehmend redaktionell erstellte Enzyklopädien
- Werden wir dies auch für Wörterbücher beobachten?
- Können wir Wörterbücher so schneller und günstiger produzieren?
- Was bedeutet das für die Qualität der Angaben?
- Was bedeutet das für die Benutzerinnen und Benutzer?
- Was kann die professionelle Lexikographie von kollaborativen Ansätzen lernen?

Auftakt

Nutzerbeteiligung in Wörterbüchern

Wörterbuchqualität

Kollaborativen Qualitätssicherung

Ausblick

- Wörterbücher beziehen ihre Nutzer immer stärker ein
- Die Grundidee ist nicht neu
 - Leseprogramme des OED seit 19. Jahrhundert (Thier, 2014)
- Aber: Erstmals werden große Mengen Wörterbuchartikel von Nutzern selbst verfasst, diskutiert und überarbeitet
- Art der Nutzerbeteiligung ist wesentlich vielfältiger geworden

**1. Direkte
Nutzerbeteiligung**

**2. Indirekte
Nutzerbeteiligung**

**3. Begleitende
Nutzerbeteiligung**

1. Direkte Nutzerbeteiligung

Anlegen, Verändern oder Löschen von Wörterbuchartikeln durch Wörterbuchbenutzer

■ Beiträge zu offen-kollaborativen Wörterbüchern

- Keine redaktionelle Kontrolle
- Artikel komplett von Nutzern verfasst



■ Beiträge zu kollaborativ-institutionellen Wörterbüchern

- Basis für verlags-/institutseigene Produkte
- Häufig keine Überarbeitung bestehender Artikel



■ Beiträge zu semi-kollaborativen Wörterbüchern

- Prüfung durch Redaktion
- Keine umfangreiche Bearbeitung



2. Indirekte Nutzerbeteiligung

Feedback für Wörterbuchanbieter

▪ Explizites Feedback

- Inhaltliche Ergänzungen (z.B. Belegstellen)
- Vorschläge zum Wörterbuch als Ganzes
- Qualitätsrückmeldungen und Korrekturen

▪ Implizites Feedback

- Häufig unbewusst erhoben
- Logdateien zur Benutzung eines Wörterbuchs
- Externe Inhalte (z.B. Flickr-Abbildungen)



3. Begleitende Nutzerbeteiligung

Austausch über die Wörterbuchinhalte hinaus

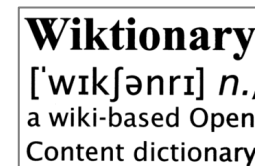
▪ Austausch zwischen Wörterbuchherstellern und -benutzern

- Blogs, Newsletter, soziale Netzwerke etc.
- Sprachspiele und wörterbuchdidaktische Angebote
- Sprachberatungsdienste



▪ Austausch der Wörterbuchbenutzer untereinander

- Foren
- Diskussionsseiten
- Nutzerkommentare



Arten der Nutzerbeteiligung

1. Direkte Nutzerbeteiligung	2. Indirekte Nutzerbeteiligung	3. Begleitende Nutzerbeteiligung
• Beiträge zu offen-kollaborativen Wörterbüchern • Beiträge zu kollaborativ-institutionellen Wörterbüchern • Beiträge zu semi-kollaborativen Wörterbüchern	• Explizites Feedback • Implizites Feedback	• Austausch zwischen Wörterbucharstellern und -benutzern – unidirektionale Kommunikation – bidirektionale Kommunikation • Austausch zwischen Wörterbuchbenutzern



Auftakt

Nutzerbeteiligung in Wörterbüchern

Wörterbuchqualität

Kollaborativen Qualitätssicherung

Ausblick

*„Wesentliches Qualitätskriterium von Wörterbüchern ist die **Zuverlässigkeit lexikographischer Daten** und deren **Auffindbarkeit**, was wiederum für den Nutzungswert von Wörterbüchern ausschlaggebend ist.“*

WLWF (2010: S. 103)

- Wörterbuchbenutzungsforschung
- Wörterbuchkritik
- Wörterbuchvergleich
- Kriterienkataloge, z.B. bei Kemmer (2010)
- Qualität der **Zwischenprodukte**
 - Auswahl der Quellen und Korpora
 - Lexikographische Instruktionen
- Qualität der **Endprodukte**
 - Printwörterbuch
 - Online-Auftritt
- Qualität des **Lexikographischen Prozesses**
 - Prüf- und Korrekturprozesse
 - Ausbildung



Sind kollaborative Wörterbücher qualitativ minderwertig?

*“Be your own lexicographer!” might be Wiktionary’s motto. Who needs experts? Why pay good money for a dictionary written by lexicographers when we can **cobble one together** ourselves?”*
—Lepore (2006: S. 87)

Wiktionary



TECHNISCHE
UNIVERSITÄT
DARMSTADT



<http://www.wordle.net> – Schriftgröße repräsentiert logarithmische Artikelanzahl

Multilingualität

Русский

Морфологические и синтаксические свойства

ло́д-ка

Существительное, неодушевлённое, женский род, 1-е склонение (тип склонения по классификации А. Зализняка).

Корень: **-лодк-**; окончание: **-а**. [Тихонов]

Произношение

МФА: [ˈlotkə]  Пример произношения

Семантические свойства

Russian

[edit]

Russische Version; russischer Eintrag
<http://ru.wiktionary.org/wiki/лодка>

Englische Version; englischer Eintrag
<http://en.wiktionary.org/wiki/boat>

Englische Version; russischer Eintrag
<http://en.wiktionary.org/wiki/лодка>

English

Etymology

From Middle English *boot*, *bot*, *boet* < **baitan* ("boat, small ship") < Proto-Norse *bēit* ("boat").

Old Norse *bātr* (whence Icelandic *bátur* & *bát*) & *bateau* are all ultimately borrowing from Latin *navis*.

Pronunciation

• Audio (US)  (file)

Noun

boat (plural **boats**)

1. A *craft* used for *transportation* of goods, fishing, racing, or *pleasure*. It may be powered by an *outboard motor* or inboard *motor* or by *wind*.
2. (*poker slang*) A full house.
3. (*chemistry*) One of two possible *conformations* of cyclohexane.

Usage notes

There's no explicit limit, but the word *boat* usually refers to

Noun

ло́дка • (lódká) f

1. (*nautical*)

Declension

declension of лодка

[show ▼]

Related terms

- лодочка, лодочник, лодчонка, подлодка
- лодочный

[edit]



Russian Wikipedia has an article on:
[Лодка](#)

[edit]

[edit]

[edit]

Wiktionary

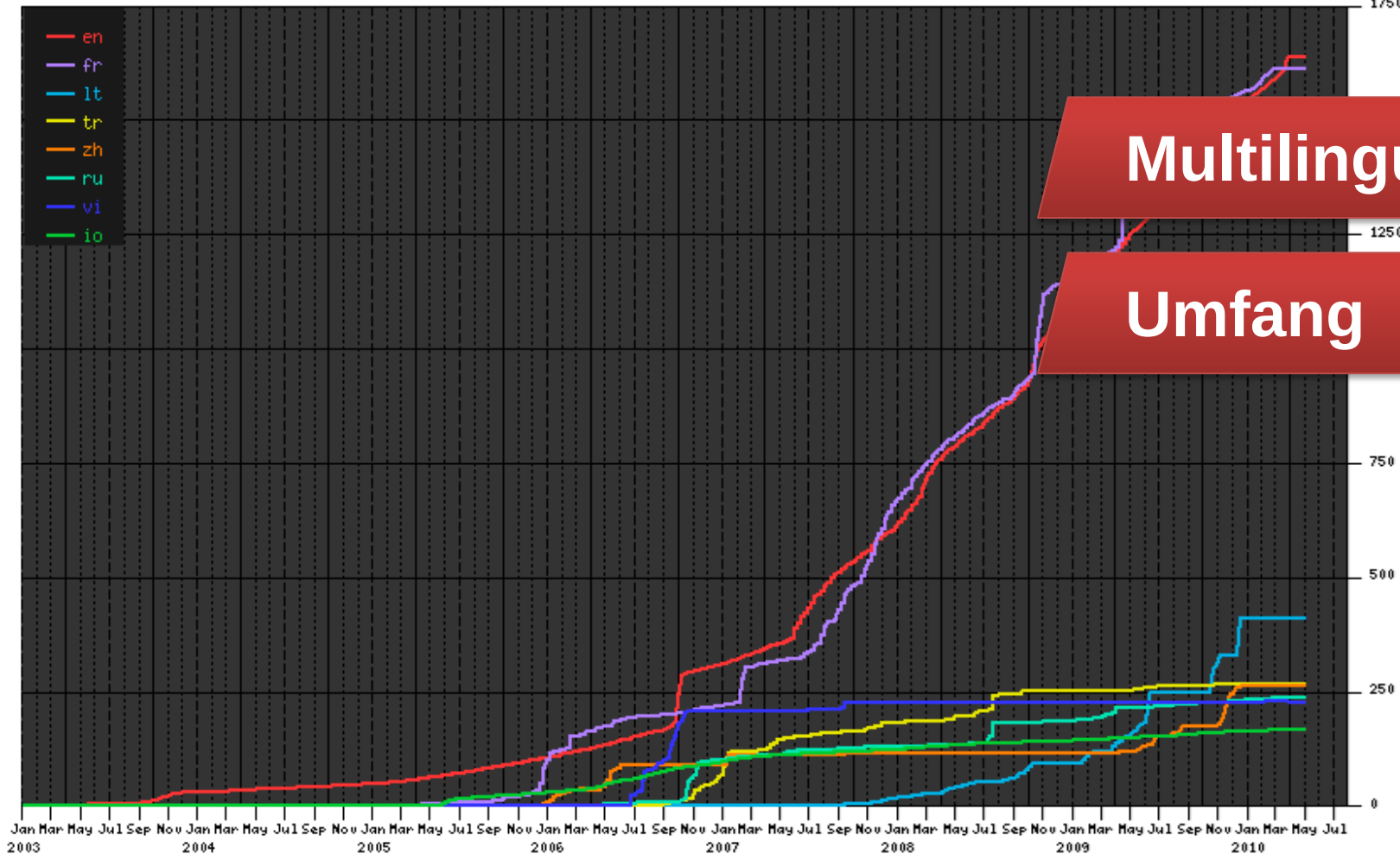


TECHNISCHE
UNIVERSITÄT
DARMSTADT

1: en - io

Article count (official)

1000 ×
1750



Multilingualität

Umfang

http://en.wikipedia.org/wiki/File:Wiktionary_growth.png

- (Unterschied | Versionen) . . winken; 16:11 . . (-2) . . 88.153.45.244 (Diskussion) (Wenn etwas in der Umgangssprache vorkommt, heißt es keinesfalls, es komme auch in der Standardsprache vor. Dementsprechend passt dort „jedoch“ nicht.)
- (Lösch-Logbuch); 16:10 . . Yoursmile (Diskussion | Beiträge) löschte Seite **пепсин** (Unsinn: Inhalt war: „na du doktr pepdine :D“; einziger Bearbeiter: 87.187.107.166)
- (Unterschied | Versionen) . . K Stabhochsprung; 16:09 . . (+2) . . 23PowerZ (Diskussion)
- (Unterschied | Versionen) . . schnorcheln; 16:02 . . (-129) . . Ыпытук (Diskussion | Beiträge) (form)
- (Lösch-Logbuch); 15:59 . . Ыпытук (Diskussion | Beiträge) löschte Seite **Pullerfrau** (oh „{{Pullerfrauen lassen sich nichts schenken}}<!--Wenn du fertig bist, lösche diese Zeile {{Sprache|Deutsch}}}} == == {{Wortart|Substa...“; einziger Bearbeiter: [[Spezial:Contributions/4...)
- (Unterschied | Versionen) . . K Villa; 15:58 . . (-11) . . Ыпытук (Diskussion | Beiträge) (Diskussion) wurden auf die letzte Version von **BetterkBot** zurückgesetzt)
- (Unterschied | Versionen) . . Wiktionary:QS/prüfe Worttrennung; 15:47 . . (+39.435) . . (aufgefüllt)
- (Unterschied | Versionen) . . K verstärken; 15:38 . . (0) . . Formatierer (Diskussion | Beiträge) (typo)
- (Unterschied | Versionen) . . Villa; 15:18 . . (+11) . . 93.199.188.170 (Diskussion) (→Villa ({{Sprache|Deutsch}}):)
- (Unterschied | Versionen) . . **Neu verschlimmern** (Konjugation); 15:18 . . (+154) . . Alexander Gamauf (Diskussion | Beiträge) (Deutsch Verbkonjugation)
- (Unterschied | Versionen) . . Flosse **[sichten]**; 15:15 . . (+31) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen:)
- (Unterschied | Versionen) . . Flosse **[sichten]**; 15:11 . . (+63) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen:)
- (Unterschied | Versionen) . . schnorcheln; 15:07 . . (+22) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen: laut w:nl:snorkelen)

Multilingualität

Umfang

Versionen

- (Unterschied | Versionen) . . winken; 16:11 . . (-2) . . 88.153.45.244 (Diskussion) (Wenn etwas in der Umgangssprache vorkommt, heißt es keinesfalls, es komme auch in der Standardsprache vor. Dementsprechend passt dort „jedoch“ nicht.)
- (Lösch-Logbuch); 16:10 . . Yoursmile (Diskussion | Beiträge) löschte Seite пенсин (Unsinn: Inhalt war: „na du doktr pepdine :D“; einziger Bearbeiter: 87.187.107.166)
- (Unterschied | Versionen) . . K Stabhochsprung; 16:09 . . (+2) . . 23PowerZ (Diskussion)
- (Unterschied | Versionen) . . schnorcheln; 16:02 . . (-129) . . Ырґаууґ (Diskussion | Beiträge) (form)
- (Lösch-Logbuch); 15:59 . . Ырґаууґ (Diskussion | Beiträge) löschte Seite Pullerfrau (oh „{{Pullerfrauen lassen sich nichts schenken}}<!--Wenn du fertig bist, lösche diese Zeile {{Sprache|Deutsch}}}} == == {{Wortart|Substa...“; einziger Bearbeiter: [[Spezial:Contributions/...)
- (Unterschied | Versionen) . . K Villa; 15:58 . . (-11) . . Ырґаууґ (Diskussion | Beiträge) (Diskussion) wurden auf die letzte Version von BetterkBot zurückgesetzt)
- (Unterschied | Versionen) . . Wiktionary:QS/prüfe Worttrennung; 15:47 . . (+39.435) . . (aufgefüllt)
- (Unterschied | Versionen) . . K verstärken; 15:38 . . (0) . . Formatierer (Diskussion | Beiträge)
- (Unterschied | Versionen) . . Villa; 15:18 . . (+11) . . 93.199.188.170 (Diskussion) (→Villa)
- (Unterschied | Versionen) . . Neu verschlimmern (Konjugation); 15:18 . . (+154) . . Ale Beiträge) (Deutsch Verbkonjugation)
- (Unterschied | Versionen) . . Flosse [sichten]; 15:15 . . (+31) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen:)
- (Unterschied | Versionen) . . Flosse [sichten]; 15:11 . . (+63) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen:)
- (Unterschied | Versionen) . . schnorcheln; 15:07 . . (+22) . . Ciciban (Diskussion | Beiträge) (→Übersetzungen: laut w:nl:snorkelen)

Multilingualität

Umfang

Versionen

Aktualität

Sind kollaborative Wörterbücher qualitativ minderwertig?

*“Be your own lexicographer!” might be Wiktionary’s motto. Who needs experts? Why pay good money for a dictionary written by lexicographers when we can **cobble one together** ourselves?”*
—Lepore (2006: S. 87)

- ✓ Multilingualität
- ✓ Umfang
- ✓ Versionen
- ✓ Aktualität

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„In the English Wiktionary, the etymologies are taken from or based on those in older dictionaries, as are the definitions, which are **extremely old-fashioned** [...]“*

—Hanks (2012: S. 77f.)

- ✓ Multilingualität
- ✓ Umfang
- ✓ Versionen
- ✓ Aktualität

Datenimport in Wiktionary

- Übernahme von Information aus anderen Sprachversionen
- Wörterbuchinhalte aus Wikipedia
- Rechtfreie Wörterbücher
- Multimedia aus freien Quellen
- Plagiarismus, z.B., aus Oxford dictionaries, Duden etc.



This page will be copied to **Wiktionary** using the automated **transwiki** process.

The information in this article appears to be suited for inclusion in a dictionary, and this article's topic meets **Wiktionary's criteria for inclusion** and will be copied into Wiktionary's transwiki space from which it can be formatted appropriately.

If this page does not meet the criteria, please remove this notice. Otherwise, the notice will be automatically removed after transwiki completes.



This page **appears to be a dictionary definition**. Please **rewrite it** to present the subject from an encyclopedic point of view. If it cannot be turned into a full encyclopedia article in the near future, consider moving it to **Wiktionary**.

Steps for moving to Wiktionary:

[\[show\]](#)

Quellenangaben in Wiktionary



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Nº	English sources	References
1)	<i>Webster's Revised Unabridged Dictionary</i> (1913)	4,088
2)	<i>The Century Dictionary and Cyclopedia</i> (1911)	2,862
3)	<i>Mineralogy Database</i> (1997 f.)	2,645
4)	<i>Wikipedia</i> (2001 f.)	2,456
5)	<i>Oxford English Dictionary</i> (multiple editions)	1,633
6)	<i>OneLook Dictionary Search</i> (1996 f.)	1,532
7)	<i>Dictionary.com</i> (1995 f.)	373
8)	<i>Online Etymology Dictionary</i> (2001 f.)	246
9)	<i>The American Heritage Dictionary of the English Language</i> (2000)	146
10)	<i>Webster's Seventh New Collegiate Dictionary</i> (1965)	122
11)	<i>American Dictionary of the English Language</i> (1828)	120
12)	<i>Random House Webster's Unabridged Dictionary</i> (1997)	108
13)	<i>Taber's Encyclopedic Medical Dictionary</i> (1993)	88
14)	<i>The New Geordie Dictionary</i> (1987)	84
15)	<i>Canadian Oxford Dictionary</i> (2004)	76

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„In the English Wiktionary, the etymologies are taken from or based on those in older dictionaries, as are the definitions, which are **extremely old-fashioned** [...]“*

—Hanks (2012: S. 77f.)

✗ **Altertümliche
Beschreibungen**
✗ **Plagiarismus**

✓ **Multilingualität**
✓ **Umfang**
✓ **Versionen**
✓ **Aktualität**

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„The Wiktionary definition is not wrong, but it is **stilted** and **archaic** in wording [...] Wikitionary's examples **do not illustrate the normal phraseology** in which the verb is used [...]“*

—Hanks (2012: S. 79)

✗ **Altertümliche
Beschreibungen**
✗ **Plagiarismus**

✓ **Multilingualität**
✓ **Umfang**
✓ **Versionen**
✓ **Aktualität**

Quellenangaben in Wiktionary

Nº	German sources	References
1)	<i>Wikipedia</i> (2001 f.)	48,732
2)	<i>Wortschatz-Lexikon</i> (1998 f.)	47,907
3)	<i>Digitales Wörterbuch der deutschen Sprache</i> (2004 f.)	45,123
4)	<i>canoonet</i> (2000 f.)	43,864
5)	<i>TheFreeDictionary.com</i> (2003 f.)	18,936
6)	<i>Deutsches Wörterbuch</i> (1854–1961)	14,021
7)	<i>Duden online</i> (2011 f.)	13,653
8)	<i>Duden: Deutsches Universalwörterbuch</i> (multiple editions)	2,859
9)	<i>Duden: Die deutsche Rechtschreibung</i> (multiple editions)	2,089
10)	<i>Duden: Das große Fremdwörterbuch</i> (multiple editions)	1,249
11)	<i>wissen.de</i> (2000 f.)	1,006
12)	<i>Metzler-Lexikon Sprache</i> (2005)	1,004
13)	<i>OWID</i> (2008 f.)	930
14)	<i>Lexikon der Sprachwissenschaft</i> (2002)	743
15)	<i>Der Neue Herder</i> (1949)	688
16)	<i>Brockhaus</i> (1996–1999)	654
17)	<i>Goethe-Wörterbuch</i> (1966 f.)	563
18)	<i>Redensarten-Index</i> (2001 f.)	541
19)	<i>Neues Lexikon der Vornamen</i> (1993)	533
20)	<i>Großes Abkürzungsbuch</i> (1980)	531

durch
Artikel-
vorlage

49% der Artikel
haben mind.
eine Quellen-
angabe

ca. 23.000
Quellen

In Wiktionary findet keine systematische Korpusanalyse statt

- Als Quellen dienen bestehende Wörterbücher, deren Angaben in Wiktionary gesammelt, gesichtet und aggregiert werden
- Zusätzlich fließen subjektive Sichtweisen und Meinungen ein
- In Diskussionsbeiträgen und durch mehrfache Überarbeitung kann dadurch Neues entstehen

Aber: die Vorgehensweise von kollaborativen Autor_innen und professionellen Lexikograph_innen ist im Regelfall sehr verschieden!

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„The Wiktionary definition is not wrong, but it is **stilted** and **archaic** in wording [...] Wiktionary's examples **do not illustrate the normal phraseology** in which the verb is used [...]“*

—Hanks (2012: S. 79)

- ✗ Altertümliche Beschreibungen
- ✗ Plagiarismus
- ✗ Korpusanalyse

- ✓ Multilingualität
- ✓ Umfang
- ✓ Versionen
- ✓ Aktualität
- ✓ Subjektivität

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„Finally these proposals have been illustrated by editing **two entries**: stock exchange and bolsa (de valores), **whose lexicographical treatment** in English Wiktionary and Spanish Wiktionary is very **deficient**“*

—Fuertes Olivera (2009: S. 130)

x Beschreibung
x Plagiarismus
x Korpusanalyse

Umfang
✓ Versionen
✓ Aktualität
✓ Subjektivität

Dauerbaustelle kollaboratives Wörterbuch

- Wörterbücher sind (in erster Linie) kein Lesebuch
- Im Regelfall ist es nicht möglich, alle Einträge im Rahmen einer Wörterbuchkritik zu prüfen
- Einschätzung erfolgt über eine Stichprobe
 - Annahme: die Qualität ist (weitgehend) stabil über das Wörterbuch hinweg, es werden gleiche Strukturen und Vorgehensweisen genutzt
- **Aber:** in kollaborativen Wörterbüchern sind Artikel in sehr unterschiedlichen Bearbeitungsstadien anzutreffen
- Heute beobachtete Qualitätsmängel sind womöglich schon morgen behoben
- Welche Artikel sind wie weit fortgeschritten?



Mit qualitativen Aussagen zu einzelnen Wörterbuchartikeln kommen wir nicht sehr weit

- Beispiel: Vergleich der Artikel *Xenolith*, *Xitsonga* und *Xylol* aus Wiktionary mit elexiko
- Ergebnis:
 - Nur Wiktionary enthält überhaupt Artikel zu *Xenolith* und *Xitsonga*
 - Beide Wörterbücher enthalten einen Artikel zu *Xylol*
 - In Wiktionary umfangreich ausgearbeitet
 - In elexiko nur wenige Kopfdaten (da noch gar nicht bearbeitet!)
- Ist elexiko nun ein qualitativ minderwertiges Wörterbuch?

**Zur Analyse kollaborativer Wörterbücher sind
qualitative und quantitative Methoden nötig!**

Quantitative Methoden benötigen vollen Zugriff auf die einzelnen Wörterbuchangaben

Redaktionelle Wörterbücher:

- Lizenzproblematik
- Proprietäre Datenformate

Kollaborative Wörterbücher:

- Mangelhafte Strukturierung der Daten
- Inkonsistenzen

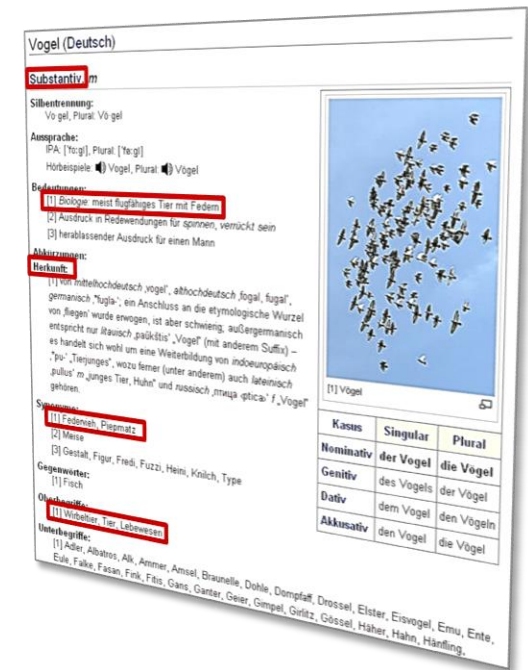
Vergleich:

- Englisch: Wiktionary – Princeton WordNet – Roget's Thesaurus
- Deutsch: Wiktionary – GermaNet – OpenThesaurus
- Russisch: Wiktionary – russisches WordNet

Funktionsumfang

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

Wiktionary
['wɪkʃənɹɪ] *n.*,
a wiki-based Open
Content dictionary



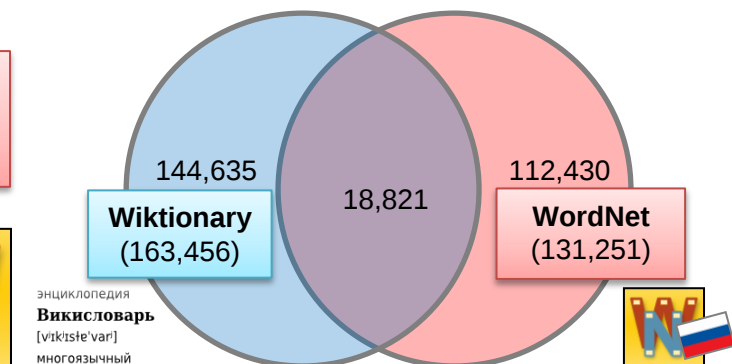
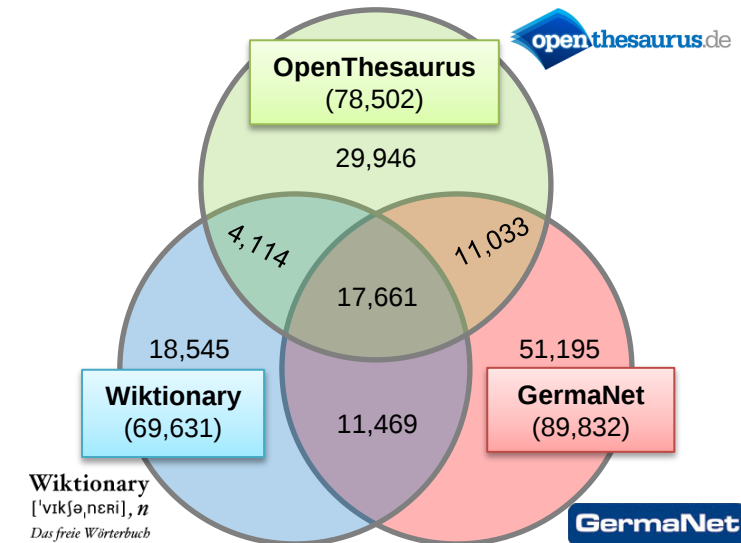
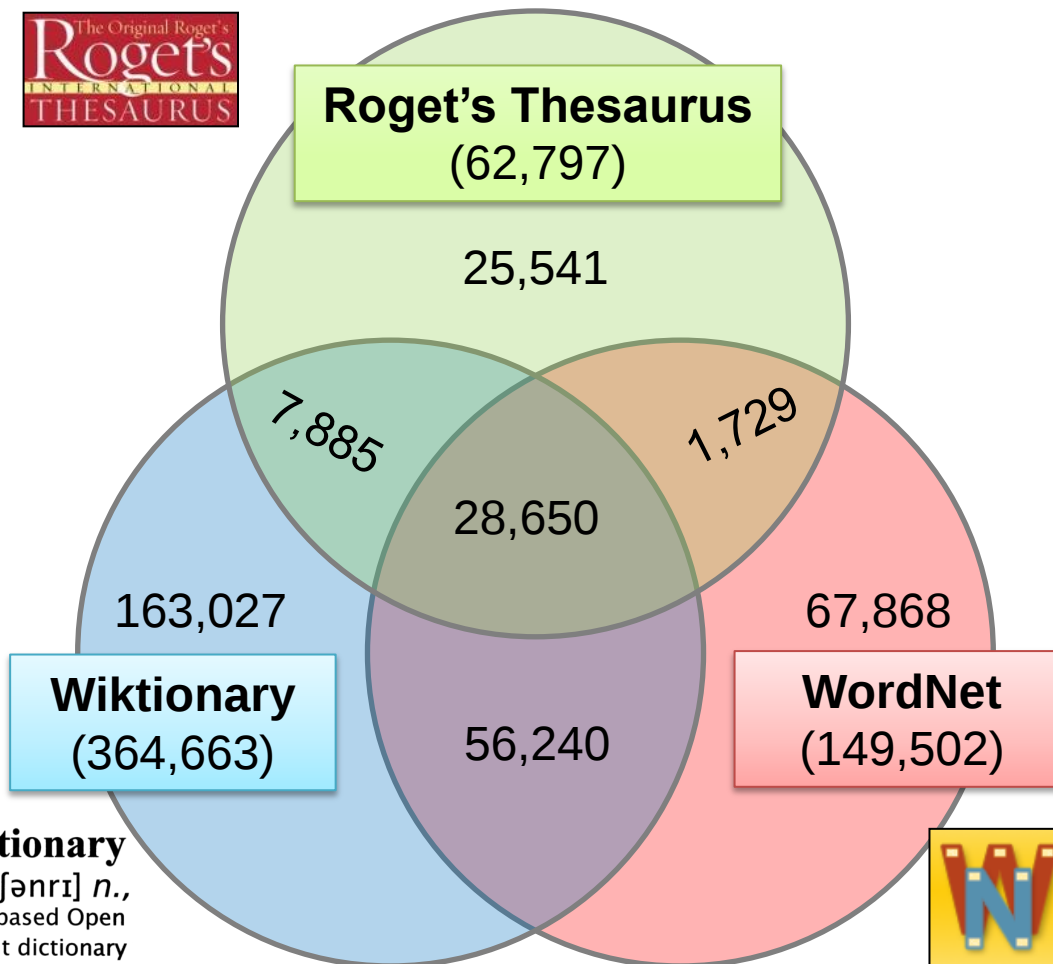
Download als quelloffene Software:

<http://www.ukp.tu-darmstadt.de/software/jwktl/>

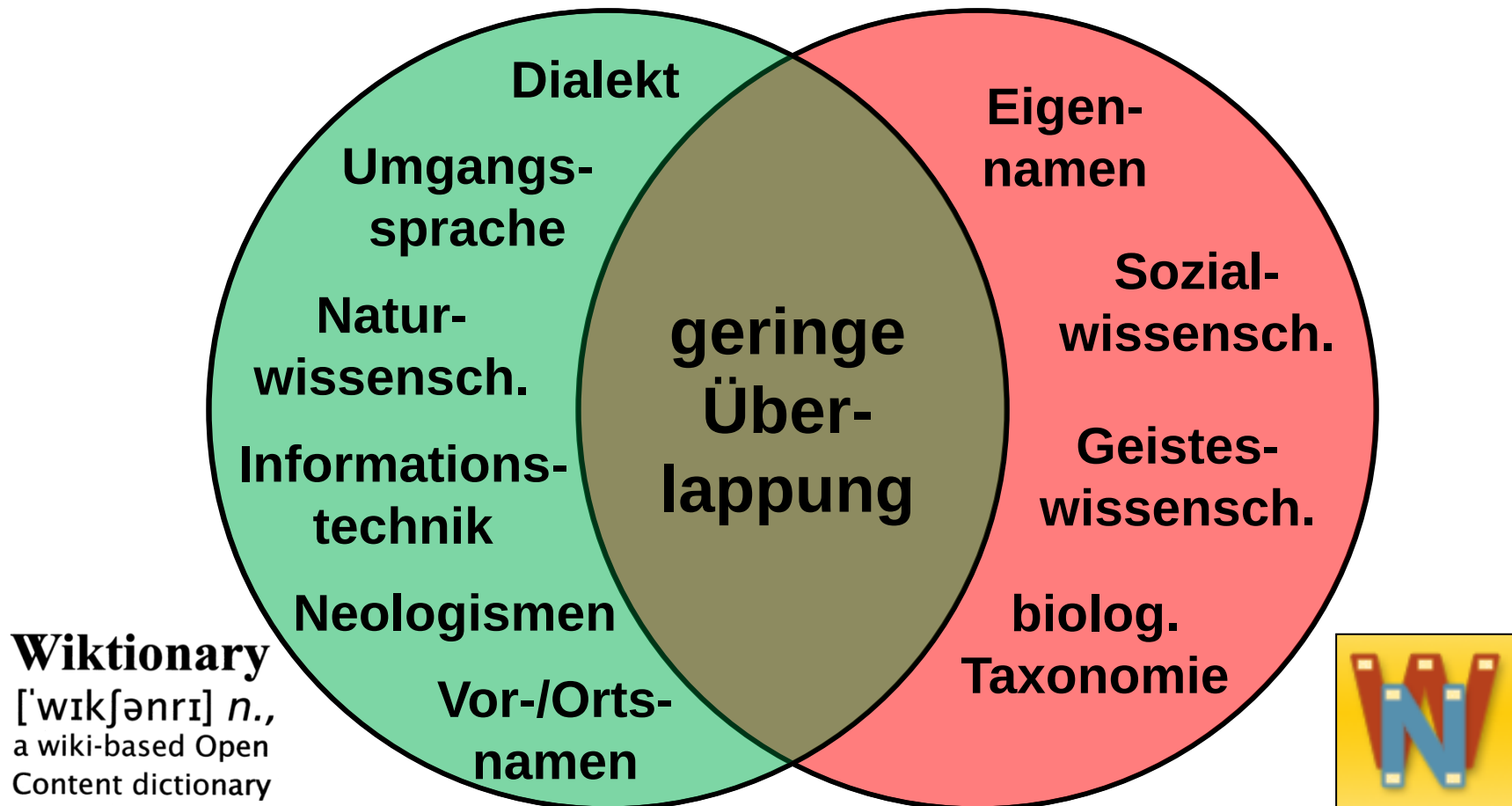
Quantitativer Wörterbuchvergleich



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Quantitativer Wörterbuchvergleich



Sind kollaborative Wörterbücher qualitativ minderwertig?

*„Finally these proposals have been illustrated by editing **two entries**: stock exchange and bolsa (de valores), **whose lexicographical treatment** in English Wiktionary and Spanish Wiktionary is very **deficient**“*
(9: S. 130)

- ✗ Altertümliche Beschreibungen
- ✗ Plagiarismus
- ✗ Korpusanalyse
- ✗ Fehlerh. Einträge
- ✗ Lücken

– Ja und Nein –

- ✓ Multilingualität
- ✓ Umfang
- ✓ Versionen
- ✓ Aktualität
- ✓ Subjektives
- ✓ Varietäten

Sind kollaborative Wörterbücher qualitativ minderwertig?

*„Aufgrund der anonymen und freiwilligen Mitarbeit kann die **Richtigkeit nie garantiert** werden. [...] Da zahlreiche Benutzer die letzten Änderungen im Auge behalten und das Wiktionary nach Fehlern durchforsten, sollten sich **Fehler in Grenzen halten.**“*
—Wiktionary (<http://de.wiktionary.org/w/index.php?oldid=1747323>)

- ✗ Plagiarismus
- ✗ Korpusanalyse
- ✗ Fehlerh. Einträge
- ✗ Lücken

– Ja und Nein –

- ✓ Versionen
- ✓ Aktualität
- ✓ Varietäten



Auftakt

Nutzerbeteiligung in Wörterbüchern

Wörterbuchqualität

Kollaborativen Qualitätssicherung

Ausblick

Maßnahmen zur Qualitätssicherung

- **Lexikographische Instruktionen**
- **Ausbildung**
- **Diskussion/Prüfung**
- **Qualitätskennzeichnungen**
 - Mängel
 - Prädikate
- **Qualitätsfeedback**

Support

1.  **Support** [Equinox](#) 23:04, 15 April 2012 (UTC)
2.  **Support.** I've had a look through the user contributions, there seem to be enough and enough evidence of progression over those 4 and a half months to merit my support. [Mglovesfun](#) (talk) 11:41, 17 April 2012 (UTC)
3.  **Support** [pidsip](#) 11:48, 17 April 2012 (UTC) looks good.

Oppose

1.  **Oppose** -[Atelaes](#) *λάλει έμοι* 23:42, 15 April 2012 (UTC) Clearly a solid editor with a good attitude, and someone destined for adminship, but they only got here in January, and I think it's just a bit too soon right now. Looking at the talk page, it looks like they're still working on some of our basic formatting. Half an admin's job is to oh so politely and patiently (or not) criticize people for outlandishly subtle and nitpicky mistakes in formatting. Also, and perhaps more importantly, I just don't have enough to go on as far as assessing their personality, though I'd like to note that what I have seen is very encouraging. However, I've seen editors who start out polite and good-natured, get promoted to admin early on (with my support vote in tow), and then turn out to be rather nasty, using their big buttons when discussion would be more appropriate. I really don't want to do that again, as bully admins are nightmarish to deal with. Everything I see points to someone whom I would support in a vote a few months down the road, but I can't do it now. Sorry. -[Atelaes](#) *λάλει έμοι* 23:42, 15 April 2012 (UTC)

I understand if you think it's too early, but as for the personality thing, I don't think it's an issue. Whether I get sysopped or not, if any admin thinks I'm being a bully, they can just give me a block until I cool down (not that I expect ever to be an angry mastodon). --[Μετáknowledge](#) *discuss/deeds* 00:45, 16 April 2012 (UTC)

Abstain

1.  **Abstain** —[Ruakh](#)*TALK* 20:26, 17 April 2012 (UTC)
2.  **Abstain** Not enough experience to judge. -- [Liliana](#) • 20:46, 17 April 2012 (UTC)

Decision

Beispiel: Kreuzung

Entschuldigt meine heftige Wortwahl, aber ich weiß nicht, was hier für "Fachleute" revertieren und sichten!!

- ein Ort, wo sich 2 Straßen treffen, nennt man "Straßenknick" oder "Straßenecke", im einfachsten Fall einfach nur Straße, wenn eine gerade verlaufende Straße von der a-Straße zur b-Straße wird.
- ein Ort, wo sich 3 Straßen treffen, wird Straßengabel oder auch Abzweigung genannt
- ein Ort, wo sich 4 oder mehr Straßen treffen, wird Kreuzung genannt. Das ist sinngleich mit der Aussage: ein Ort, wo sich 2 (oder mehr) Straßen kreuzen *oder* ein Ort, wo eine Straße eine zweite Straße quert.

Ich weiß wirklich nicht, was ihr hier den Lesern für Bedeutungen "unterjubeln" wollt, aber das ist schon grenzwertiges Verhalten. Da ich bei einer massiven Revertierung ohne jede Rückfrage erheblichen Sachverstand erwarte bzw. unterstelle, kann ich hier nur VM annehmen.

Ich möchte jetzt hier keinen Editwar anzetteln, weshalb ich die stattgefunden Revertierung meiner Änderungen nicht erneut ändere und bitte, jemand mit Sachkunde und Verständnis möge sich der Angelegenheit annehmen. —[JÄhh \(Diskussion\)](#) 12:03, 7. Mär. 2011 (MEZ)

Gerne. Eine Straße, die in eine Kreuzung führt, endet nicht dort, sondern läuft, salopp gesagt, einfach durch. So kommt man auf zwei oder mehr sich treffende Straßen. Definitorisch ebenso erfasst im [DWDS](#) , in *Duden Das große Wörterbuch der deutschen Sprache in 10 Bänden. 3., völlig neu bearbeitete und erweiterte Auflage. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 1999*, unter englisch "crossroads" im *Oxford Dictionary of English* und bei [Merriam-Webster](#) . Wo steht es anders? Davon abgesehen ist dein Ton wieder einmal völlig unangemessen. —[Pill \(Kontakt\)](#) 12:50, 7. Mär. 2011 (MEZ)
Oder sehe ich jetzt das Problem nicht? Du siehst einen Unterschied zwischen "treffen" und "kreuzen"? —[Pill \(Kontakt\)](#)

Baustellenkennzeichnung: Überarbeiten

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

Dieser Eintrag oder Abschnitt bedarf einer Überarbeitung. Hilf bitte mit, ihn zu verbessern, und entferne anschließend diese Markierung.



Folgendes ist zu überarbeiten: Referenzen zuordnen; Bedeutungen abgrenzen; Gegenworte prüfen; Achsen-, Dreh-, Punkt-Sym./"symmetrisch"; erweitern; ...

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

Worttrennung:

Sym·me·trie, Plural: Sym·me·trien

Aussprache:

IPA: [zɪmɐ'tri:], Plural: [zɪmɐ'tri:ən]

Hörbeispiele: —, Plural: —

Bedeutungen:

[1] spiegelbildliche Gleichheit

[2] *Mathematik*: eine Eigenschaft von Relationen

Herkunft:

von lateinisch **symmetria** ^{→ la}, das auf griechisch **συμμετρία** (symmetria) ^{→ grc} zurückgeht. Zugrunde liegen griechisch **συν** (syn) ^{→ grc} (syn, vergleiche **syn-**) „gemeinsam“ und griechisch **μέτρον** (métron) ^{→ grc} „(das rechte) Maß, Länge, Größe, auch Silben- oder Versmaß“.^[1]

Kasus	Singular	Plural
Nominativ	die Symmetrie	die Symmetrien
Genitiv	der Symmetrie	der Symmetrien
Dativ	der Symmetrie	den Symmetrien
Akkusativ	die Symmetrie	die Symmetrien

Baustellenkennzeichnung: Löschkandidat



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Dieser Eintrag oder Teile davon wurden zur Löschung vorgeschlagen. Er erfüllt vermutlich die Qualitätskriterien für das Wiktionary nicht. Wenn du der Meinung bist, dass dieser Eintrag nicht gelöscht werden sollte, so kannst du dich an der Diskussion auf der Seite der [Löschkandidaten](#) beteiligen. Wenn er dort noch nicht eingetragen ist, dann holst du das am besten nach. Über die Löschung wird nach sieben Tagen entschieden. Während dieser Zeit ist es natürlich möglich, den Eintrag zu erweitern und zu verbessern. Halte dich dabei aber bitte an die [Formatvorlage](#).

Hier der **konkrete Grund**, warum der Eintrag gelöscht werden sollte: Keine Belege, ganze [78 Google-Treffer](#).
—/KRONf/ (Diskussion) 16:48, 13. Okt. 2011 (MESZ)

ausschaltbereit ([Deutsch](#)) [[Bearbeiten](#)]

Adjektiv [[Bearbeiten](#)]

Worttrennung:

aus·schalt·be·reit, *keine Steigerung*

Aussprache:

IPA: [ˈaʊ̯ʃaltbɐˈrɛɪt]

Hörbeispiele: —

Bedeutungen:

[1] [bereit zum Ausschalten](#)

Positiv	Komparativ	Superlativ
ausschaltbereit	—	—
Alle weiteren Formen: ausschaltbereit (Deklination)		

Stabilversionen

[unmarkierte Version]

[gesichtete Version]

Version vom 29. November 2009, 17:21 Uhr
(bearbeiten)

88.74.55.111 (Diskussion)

← Zum vorherigen Versionsunterschied

Version vom 29. November 2009, 18:14 Uhr
(bearbeiten) (rückgängig machen)

Stepro (Diskussion | Beiträge)

(größte Fehler beseitigt, Bedeutungen erweitert,
Referenzen zugeordnet, Beispiele ergänzt)

Zum nächsten Versionsunterschied →

Version A

Das Ziel der gesichteten Versionen im Wiktionary ist, dem Leser eine benutzbare Seite zu liefern. → frei von Vandalismus

Version B

Das Ziel der gesichteten Versionen im Wiktionary ist, dem Leser eine mögliche hohe Qualität zu liefern. → frei von Vandalismus und inhaltlich korrekt

Is this translation correct and useful?	
English	loodering [Scot.] [Irish]
German	heftige Prügelei {f}
Class	noun
Comment	• http://www.oed.com/view/Entry/... " 'loodering n. a severe beating; a hiding. 1866 W. Gregor Dial. Banffshire (Philol. Soc.) 227 Llouthera severe beating. 1912 J. Campbell Judgment i. 9 She'll get over it. It's not the 1985 L. Shannon in Stories (Boston) No. 12 14 Whenever s biggest loodering she's ever had in her whole life. "
	YES (100% correct) NO / MAYBE ...



Auftakt

Nutzerbeteiligung in Wörterbüchern

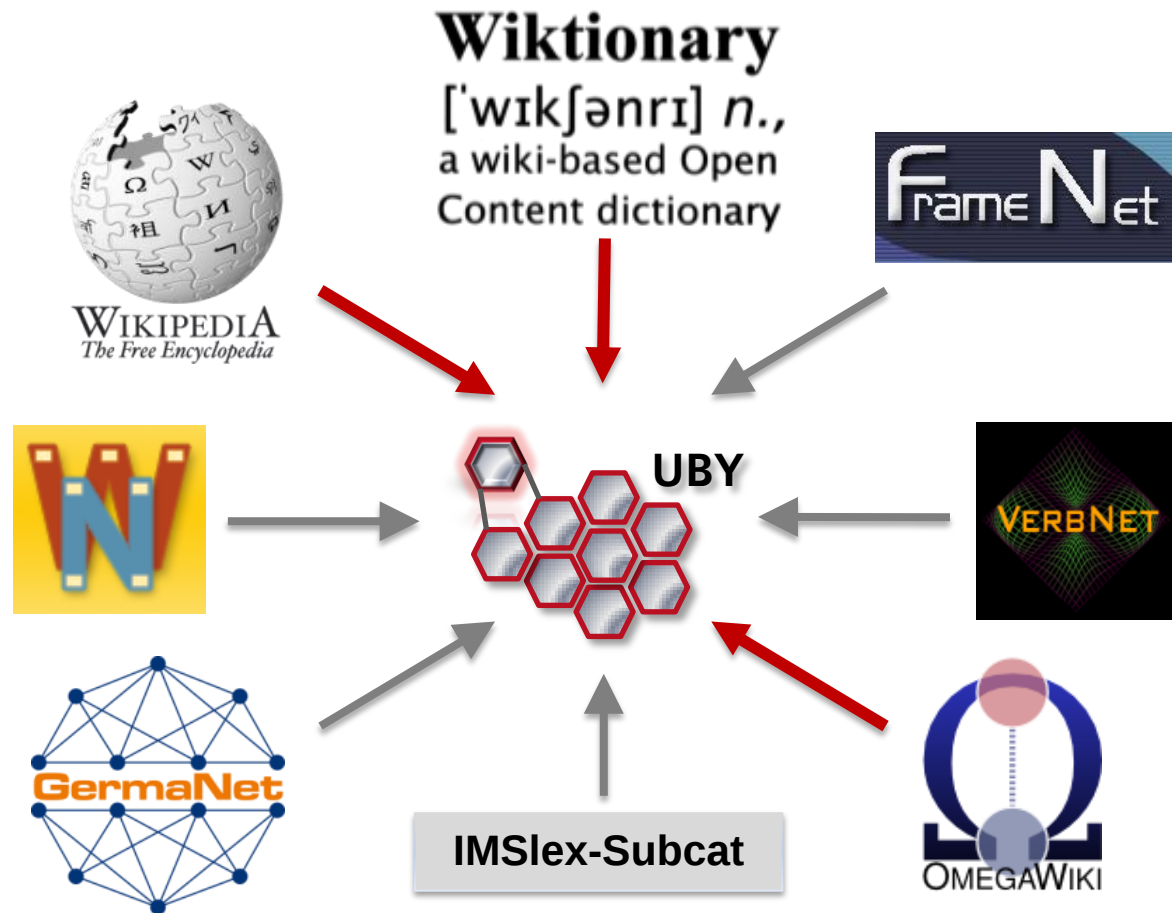
Wörterbuchqualität

Kollaborativen Qualitätssicherung

Ausblick

Quo vadis, Lexikographie?

- **Kollaborative Wörterbücher sind kein adäquater Ersatz für redaktionelle Wörterbücher**
 - Zahlreiche Qualitätsmängel
 - Fehlende Korpusanalyse
- **Der kollaborative Ansatz bietet jedoch großen Mehrwert**
 - Fachsprachen, Dialekte und Umgangssprache
 - Schneller Ausbau, zeitnahe Aktualisierung
 - Chance für kleinere Sprachgemeinden, umfangreiche Übersetzungen
 - Wörterbuchdidaktik: gute Laienautoren sind gute Wörterbuchbenutzer!
- **Redaktionelle Wörterbücher stehen vor großen ökonomischen Herausforderungen**
 - Kooperationsmodelle zwischen freiwilligen Nutzer_innen und Experten?
 - Kombination von kollaborativ erstellten und redaktionell erstellten Inhalten?



<https://www.ukp.tu-darmstadt.de/data/lexical-resources/uby/>



Hand- und Studienbuch „Internetlexikografie“ (derzeit in Vorbereitung)

- Typologie von Internetwörterbüchern und -portalen
- Der lexikografische Prozess
- Datenmodellierung
- Vernetzungsstrategien und Zugriffsstrukturen
- Automatische Gewinnung von lexikografischen Angaben
- Benutzeroberfläche von Internetwörterbüchern
- Nutzerbeteiligung
- Wörterbuchbenutzungsforschung

<http://www.internetlexikografie.de/>

Weiterführende Literatur

- A. Abel, C.M. Meyer: **Nutzerbeteiligung**, Kapitel 8 in A. Klosa (Hrsg.): *Handbuch Internetlexikografie*, Berlin: Akademie Verlag, in Vorbereitung.
- C.M. Meyer, I. Gurevych: **Methoden bei kollaborativen Wörterbüchern**, *Lexicographica* 30(1): 187–212, 2014.
- C.M. Meyer: **Wiktionary: The Metalexicographic and the Natural Language Processing Perspective** (= tprints 3654). Dissertation, Darmstadt: Technische Universität Darmstadt 2013.
<http://tprints.ulb.tu-darmstadt.de/3654/>
- A. Abel, C.M. Meyer: **The dynamics outside the paper: user contributions to online dictionaries**, in: *Proceedings of the 3rd eLex conference 'Electronic lexicography in the 21st century: thinking outside the paper'*, S. 179–194, 2013. Tallinn, Estland.

Vielen Dank für die Aufmerksamkeit!
Fragen/Anregungen?

- M. Matuschek, C.M. Meyer, I. Gurevych: **Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Translation Applications**, *Translation: Computation, Corpora, Cognition – Special Issue "Language Technology for a Multilingual Europe"* 3(1):87–118, 2013.
- I. Gurevych, J. Eckle-Kohler, S. Hartmann, M. Matuschek, C.M. Meyer, C. Wirth: **UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF**, in: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, p. 580–590, April 2012. Avignon, France.
- C.M. Meyer, I. Gurevych: **Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography**. Kapitel 13 in S. Granger, M. Paquot (Hrsg.): *Electronic Lexicography*, S. 259–291, Oxford: Oxford University Press, November 2012.

- P.A. Fuertes-Olivera: **The Function Theory of Lexicography and Electronic Dictionaries: Wiktionary as a Prototype of Collective Free Multiple-Language Internet Dictionary**, in H. Bergenholtz, S. Nielsen, S. Tarp (Hrsg.): *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*, S. 99–134, Bern: Peter Lang, 2009.
- P. Hanks: **Word Meaning and Word Use: Corpus evidence and electronic lexicography**, in S. Granger, M. Paquot (Hrsg.): *Electronic Lexicography*, S. 57–82, Oxford: Oxford University Press, 2012.
- K. Kemmer: **Onlinewörterbücher in der Wörterbuchkritik. Ein Evaluationsraster mit 39 Beurteilungskriterien** (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2010), Mannheim: Institut für Deutsche Sprache, 2010.
- J. Lepore: **Noah's Mark: Webster and the original dictionary wars**. *The New Yorker* LXXXII(36): 78–87. New York, NY: Condé Nast Publications, 2006.
- J.P. Rohweder, G. Kasten, D. Malzahn, A. Piro, J. Schmid: **Informationsqualität – Definitionen, Dimensionen und Begriffe**, Kapitel 2 in K. Hildebrand, M. Gebauer, H. Hinrichs, M. Mielke (Hrsg.): *Daten- und Informationsqualität*, S. 25–45, Wiesbaden: Vieweg+Teubner, 2008.
- K. Thier: **Das Oxford English Dictionary und seine Nutzer**, in A. Abel, A. Klosa (Hrsg.): *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“* (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014), S. 62–69, Mannheim: Institut für Deutsche Sprache, 2014.
- R.Y. Wang, D.M. Strong: **Beyond Accuracy: What Data Quality Means to Data Consumers**, *Journal of Management Information Systems* 12(4): 5–34, 1996.
- WLWF = H.E. Wiegand, M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer, W. Wolski (Hrsg.): **Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research**, Vol. 1 (Systematische Einführung / Systematic Introduction, A–C), Berlin/New York: de Gruyter, 2010.



Kontakt / Contact

Christian M. Meyer

Technische Universität Darmstadt
Ubiquitous Knowledge Processing Lab

📍 Hochschulstr. 10, 64289 Darmstadt, Germany

☎ +49 (0)6151 16–5386

📠 +49 (0)6151 16–5455

✉ 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.