



Data Lexsikografis

Pengantar menuju Leksem di Wikidata

Raisha Abdillah
Project Lead,
Wikidata Software Collaboration

Kami merekam sesi ini. Harap senyapkan mikrofon dan kamera Anda saat Anda tidak sedang berbicara.

Agenda

- Wikidata: Rekap singkat
- Kata dalam Wikidata: Bagaimana cara kerjanya?
- Mengapa data leksikografis di Wikidata itu penting?
- Apa yang dapat kita lakukan dengan Leksem?
- Kondisi data leksikografis di Wikidata
- Pranala penting



Wikidata: Rekap singkat

Tentang Wikidata



- Basis pengetahuan
- Bagian dari proyek Wikimedia
- Data terstruktur
- Bertaut ke basis data lainnya
- Multibahasa
- Kolaboratif
- Dirilis dengan lisensi domain publik (CC0)
- Berdasarkan fakta dan referensi
- Dibuat untuk manusia dan mesin

Butir (*Item*) di Wikidata



- Halaman utama
- Portal komunitas
- Warung Kopi
- Buat butir baru
- Perubahan terbaru
- Butir sembarang
- Layanan Kueni Sekitaran
- Bantuan
- Menyumbang
- Data leksikografi
- Buat leksem baru
- Perubahan terbaru
- Leksem sembarang
- Perkakas
- Pranala balik
- Perubahan terkait
- Halaman istimewa
- Pranala permanen
- Informasi halaman
- Konsep Pengidentifikasi
- Sumber Seragam (URI)
- Kutip halaman ini
- Lihat URL pendek
- Unduh kode QR
- Cek interwiki

Halaman **Pembicaraan**

Bahasa Indonesia Raisha (WSC) Pembicaraan Preferensi Beta Daftar pantauan Kontribusi Keluar log

Baca [Daftar label](#) [Lihat riwayat](#) [Lainnya](#)

Kartini (Q1373764)***

pahlawan nasional Indonesia (1879-1904)

R. A. Kartini | Raden Adjeng Kartini | Raden Ajeng Kartini | Raden Ayu Kartini | Kanjeng Raden Ayu Adipati Djojo Hadiningrat, | Raden Ayu Adipati Kartini Djojoadhinigrat

Dalam bahasa lainnya

Bahasa	Label	Deskripsi	Juga dikenal sebagai
Bahasa Indonesia	Kartini	pahlawan nasional Indonesia (1879-1904)	R. A. Kartini Raden Adjeng Kartini Raden Ajeng Kartini Raden Ayu Kartini Kanjeng Raden Ayu Adipati Djoj... Raden Ayu Adipati Kartini Djojo...
Inggris	Kartini	Indonesian heroine, feminist	Raden Adjeng Kartini R. A. Kartini Raden Ayu Adiningrat R. A. Kartini Djojo Adhinigrat Raden Ajeng Kartini Kanjeng Raden Ayu Adipati Djoj...
Turki	Gusti Raden Ayu Adipati Kartini Djojoadiningrat,	Tidak ada deskripsi	Raden Adjeng Kartini Kartini


Semua bahasa yang dimasukkan

Pernyataan

adalah manusia *** [sunting](#)

[1 rujukan](#)

[+ tambah nilai](#)

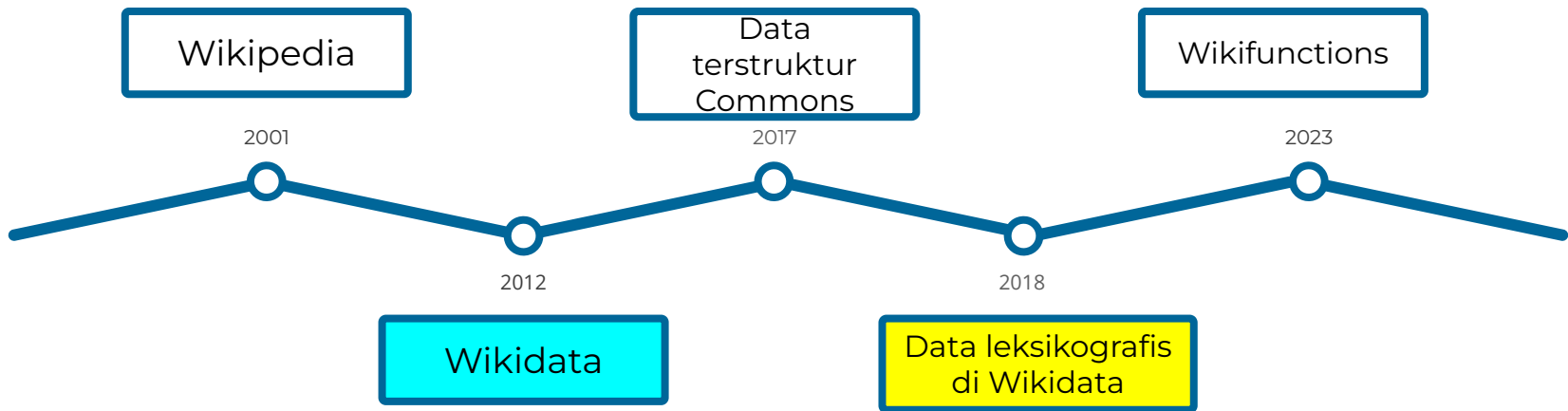
gambar  [sunting](#)

Wikipedia (34 entri) [sunting](#)

- ar كارتيڤي
- ban Kartini
- bbc Kartini
- bjn Kartini
- ca Kartini
- cy Kartini
- de Kartini
- en Kartini
- es Kartini
- eu Kartini
- fi Raden Ajeng Kartini
- fr Kartini
- gl Kartini
- gor Kartini
- hy Կարտինի (հասարակական գործիչ)
- id Kartini
- it Kartini
- ja カルティニ
- lv Kartini
- ko 카르티니
- mi Կարտինի
- ms R. A. Kartini
- nl Kartini
- pl Kartini
- ru Картина
- su Kartini
- sv Kartini
- ta கார்தினி
- th ฅารทิณี
- tr Raden Adjeng Kartini
- uk Картіні



Kata dalam Wikidata: Bagaimana cara kerjanya?



Wikidata @wikidata · 11h
First experiment of lexicographical data is out! You can now describe words and phrases in your language in Wikidata lo/ wikidata.org/wiki/Wikidata... #LexData

9 replies 53 retweets 58 likes

Esperanto LiberaScio @ELiberaScio
Replying to @wikidata
Many thanks. This will help small languages a LOT!

10:24 PM · 23 May 2018

Bedanya apa, ya?

“tikus” sebagai konsep

- Spesies mamalia
- Nama takson
- Ukuran rata-rata
- Gambar
- ID Encyclopedia of Life



File:House mouse.jpg,
public domain

“tikus” sebagai Leksem

- Bahasa: Indonesia
- Kelas kata: kata benda (nomina)
- Bentuk jamak: tikus-tikus
- Makna: binatang
- Terjemahan: mouse (en), rato (pt)
- Pemenggalan: ti·kus

Glosarium:

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Glossary/id

Model data:

[MediaWiki:Extension:WikibaseLexeme/Data Model](https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Model)





(L1330761)

tikus

id




Language Indonesia

Lexical category nomina

Pernyataan

contoh penggunaan



Ada seekor tikus tinggal di lobang bambu yang dipinggir jurang itu 



(Bahasa Indonesia) •••

bentuk subjek

tikus

makna subjek

tikus (Indonesia) - binatang pengerat, umumnya berwarna hitam, abu-abu, atau putih

▶ 1 rujukan

+ tambah nilai

asal kata



tikus/تيكوس •••







▼ 0 rujukan


+ tambah rujukan

- Halaman utama
- Portal komunitas
- Warung Kopi
- Buat butir baru
- Perubahan terbaru
- Butir sembarang
- Layanan Kueri
- Sekitaran
- Bantuan
- Menyumbang

- Data leksikografi
- Buat leksem baru
- Perubahan terbaru
- Leksem sembarang

- Perkakas
- Pranala balik
- Perubahan terkait
- Halaman istimewa
- Pranala permanen
- Informasi halaman
- Konsep Pengidentifikasi Sumber Seragam (URI)
- Kutip halaman ini

Senses

L1330761-S1	Bahasa Indonesia	binatang pengerat, umumnya berwarna hitam, abu-abu, atau putih	 sunting
	Inggris	mouse	
	Jerman	Maus	
	Turki	fare	
	Jepang	ねずみ	

Pernyataan tentang L1330761-S1




[gambar](#)    [sunting](#) 

Apodemus sylvaticus bosmuis.jpg
2.160 × 1.524; 2,53 MB

▼ 0 rujukan

[+ tambah rujukan](#)

[+ tambah nilai](#)

[butir untuk makna ini](#)  [tikus](#)  [sunting](#) 

▼ 0 rujukan

[+ tambah rujukan](#)

[+ tambah nilai](#)

Forms

L1330761-F1

tikus
id

 [sunting](#)

Grammatical features [tunggal](#)

Pernyataan tentang L1330761-F1

[audio pengucapan](#)  





▶ 0:00 / 0:00  

[LL-Q9240 \(ind\)-Fexpr-tikus.wav](#)
1,0 d; 94 KB

▼ 0 rujukan

[+ tambah rujukan](#)

[+ tambah nilai](#)

[pemenggalan suku kata](#)  

[ti:kus](#)

▼ 0 rujukan

[+ tambah rujukan](#)

[+ tambah nilai](#)

[+ tambah pernyataan](#)

L1330761-F2

tikus-tikus
id

 [sunting](#)

Grammatical features [jamak](#)

Pernyataan tentang L1330761-F2

[+ tambah pernyataan](#)



Mengapa data leksikografis di Wikidata itu penting?

Mengapa ini menarik?

- Data terstruktur = dapat dibaca oleh mesin
- Dapat digunakan kembali oleh perangkat lunak, penelitian, kamus, layanan penerjemahan
- CC0 = pengetahuan terbuka, dapat digunakan kembali oleh semua orang
- Berbagai macam bahasa, termasuk bahasa yang kurang terwakili
- Komunitas internasional = lebih banyak orang yang dapat membantu

Apa perbedaannya dengan Wiktionary?

- Wiktionary = teks polos + templat, Wikidata = data terstruktur
- Wikidata dapat dengan mudah diurai dan digunakan kembali
- Wikidata bekerja dengan Leksem, Wiktionary menggabungkan Leksem yang berbeda dalam artikel yang sama
- Wikidata = CC0, Wiktionary = CC-BY SA
- Wikidata dapat digunakan untuk mendukung Wiktionary (apabila komunitas setuju)

Apa perbedaannya dengan layanan lain?

- Kami menyediakan data yang dapat dimanfaatkan untuk pembuatan perangkat lunak
- Kami melakukan lebih dari sekadar penerjemahan: kami membantu mesin memahami bahasa
- Kami memberikan akses ke data dengan lisensi CC0
- Kami menyertakan semua bahasa, tidak hanya bahasa yang paling menguntungkan
- Kami memberdayakan orang-orang untuk berkontribusi pada data



Apa yang dapat kita lakukan dengan Leksem?

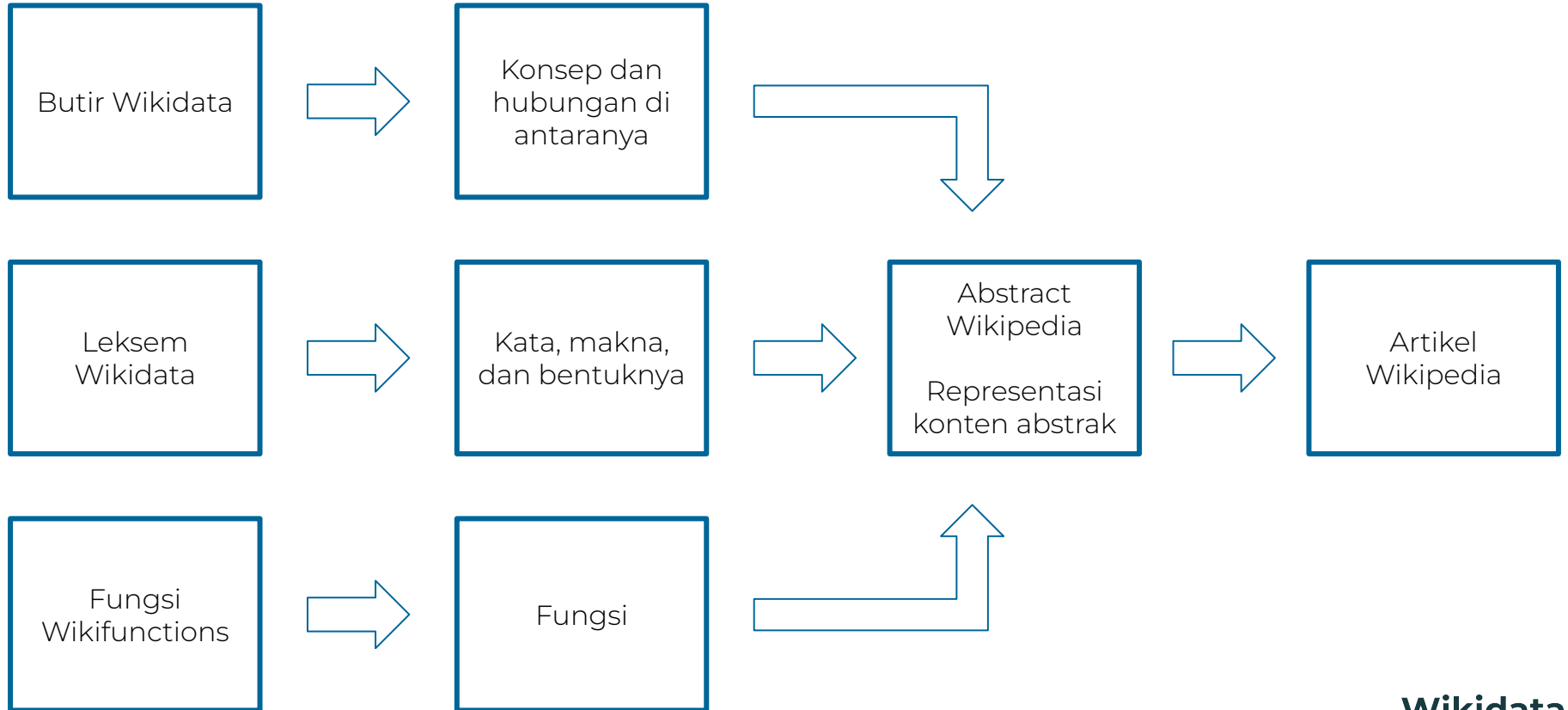
Mendukung proyek Wikimedia

- Menyediakan data terstruktur yang dapat digunakan kembali pada halaman Wiki
- Memungkinkan kontributor untuk menyunting data yang sama
- Pembuatan perangkat baru yang memudahkan kontribusi
- Pembuatan teks, misalnya untuk Abstract Wikipedia

Abstract Wikipedia

- Pembuatan teks yang dapat dibaca manusia untuk Wikipedia
- ... dalam berbagai bahasa
- Berdasarkan data yang tersedia

Komponen Abstract Wikipedia




```
Article(
  content: [
    Instantiation(
      instance: San Francisco (Q62),
      class: Object_with_modifier_and_of(
        object: center,
        modifier: And_modifier(
          conjuncts: [cultural, commercial, financial]
        ),
        of: Northern California (Q1066807)
      )
    ),
    Ranking(
      subject: San Francisco (Q62),
      rank: 4,
      object: city (Q515),
      by: population (Q1613416),
      local_constraint: California (Q99),
      after: [Los Angeles (Q65), San Diego (Q16552), San Jose (Q16553)]
    )
  ]
)
```



San Francisco is the cultural, commercial, and financial center of Northern California. It is the fourth-most populous city in California, after Los Angeles, San Diego and San Jose.



San Francisco ist das kulturelle, kommerzielle und finanzielle Zentrum Nordkaliforniens. Es ist, nach Los Angeles, San Diego und San Jose, die viertgrößte Stadt in Kalifornien.



...

Pemanfaatan sebagai kamus

- Mencari definisi dan terjemahan
- Kamus dengan tujuan khusus (rima, topik tertentu)
- Tesaurus dan kamus sinonim
- Membangun alat penerjemahan (terutama untuk bahasa yang belum memiliki alat penerjemahan)

The Surrounding Ocean

Animalia – English, proper noun
animal – English, noun

mourningly – English, adverb

mousepad – English, noun
マウス – Japanese, noun

Maus – Bavarian, noun – small gray or brown mammal with predominantly long tail, rat-like, in the breeding form also white or black

mouse – English, noun
1. any small rodent of the genus *Mus*
2. input device



jengbariga – Dagbani, noun – any small rodent of the genus *Mus*

mouselike – English, adverb

Navigation icons: left arrow, right arrow, search, and a red X icon.

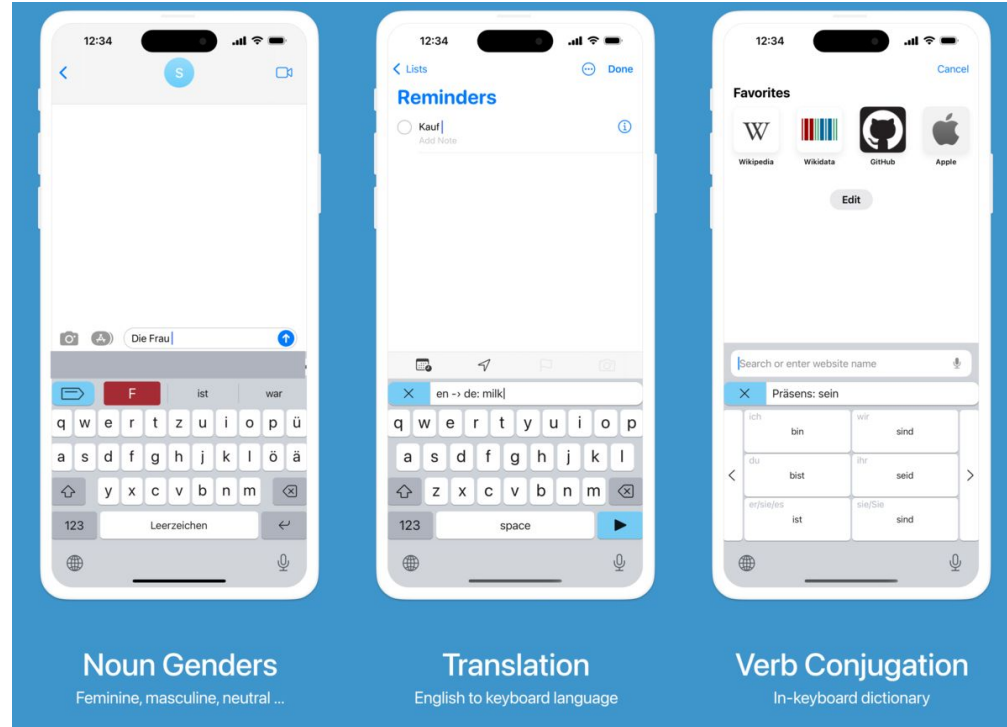
Jelajahi kata dan artinya, terjemahan, dan sinonimnya: [d:Wikidata:The Surrounding Ocean](https://www.wikidata.org/wiki/d:Wikidata:The_Surrounding_Ocean)

Papan ketik yang didukung Wikidata

- Menyediakan terjemahan saat mengetik dalam bahasa kedua
- Membantu penulisan pada bahasa dengan gender gramatikal, konjugasi kata kerja, dan penggunaan kata depan
- Anotasi kata

Papan ketik Scribe

- Papan ketik untuk orang yang sedang mempelajari bahasa kedua
- Saat ini tersedia dalam bahasa Prancis, Jerman, Italia, Portugis, Rusia, Spanyol, Swedia (bahasa lainnya akan ditambahkan)



<https://github.com/scribe-org>

**Wikidata
Lexicodays**
28-30 June 2024

Alat pembelajaran bahasa lainnya

- Membuat daftar kata dan pelajaran
- Mengilustrasikan kata-kata
- Membuat permainan dan latihan

Riset

- Bagaimana bahasa berevolusi dari waktu ke waktu, kelas sosial, dan lainnya?
- Apakah kelas kata berubah maknanya dari waktu ke waktu?
- Melokalkan kata-kata di peta

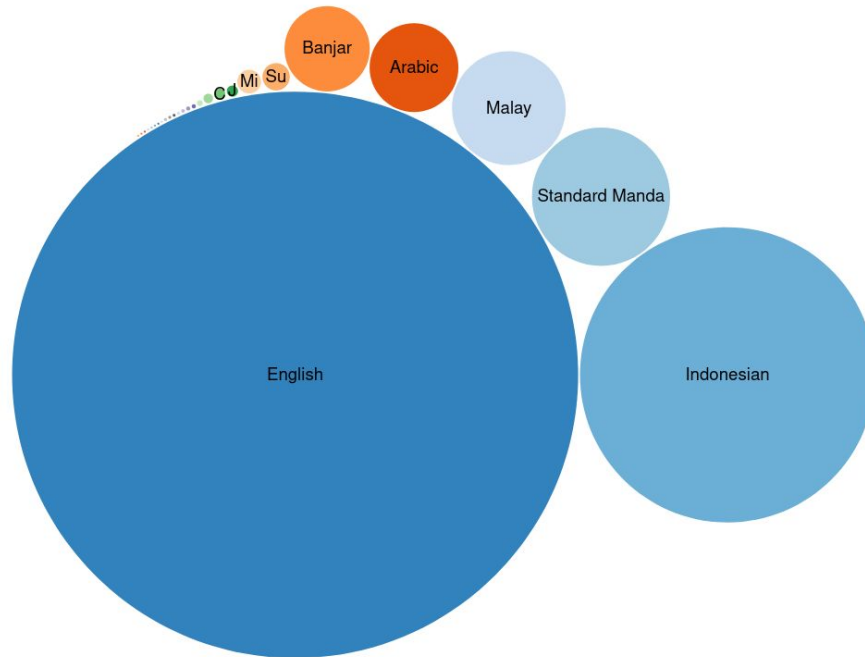
Analisis teks

- Analisis sentimen
- Penandaan untuk kelas kata
- Pengenalan entitas bernama



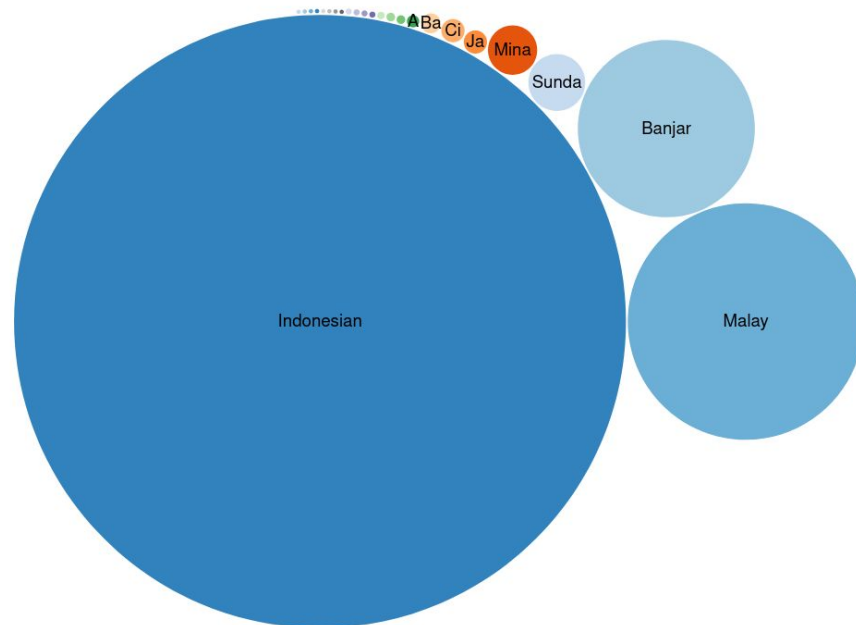
Kondisi data leksikografis di Wikidata

Leksem bahasa Indonesia di Wikidata Leksem



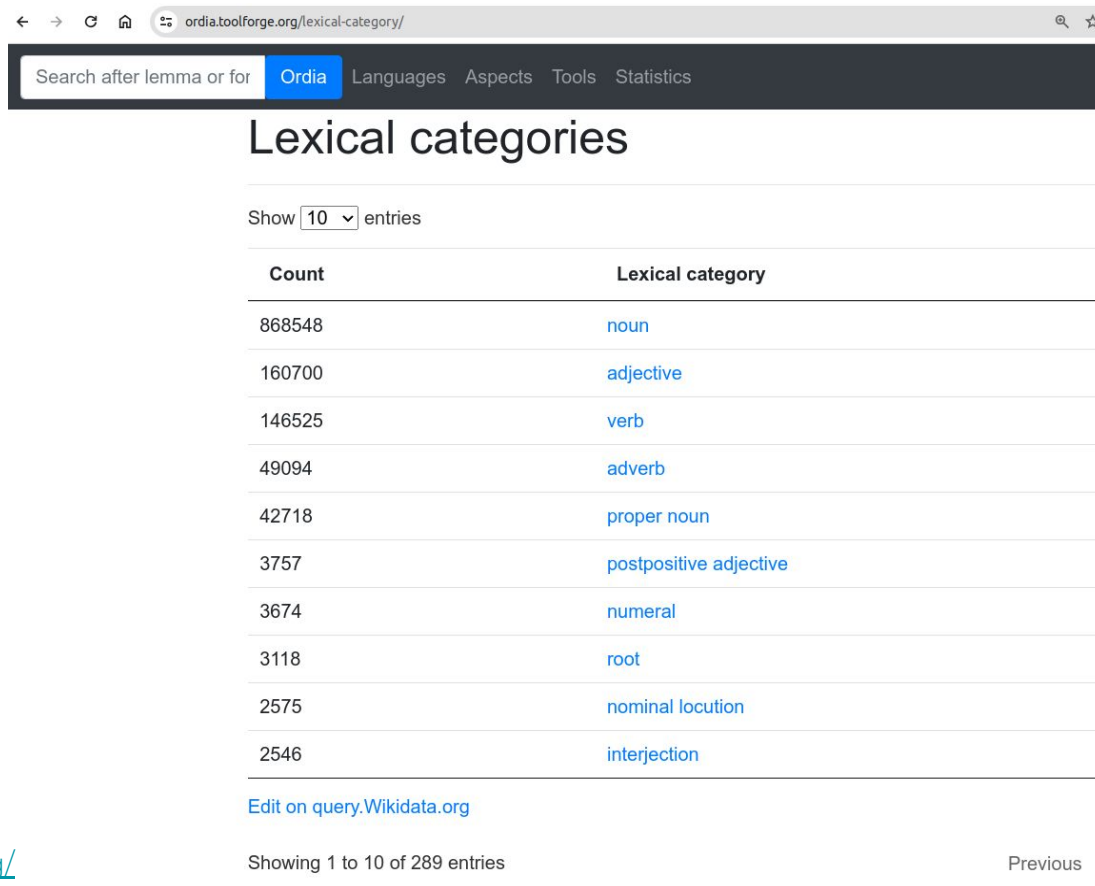
Coba: <https://w.wiki/AJ5U>

Leksem bahasa yang dituturkan di Indonesia (di luar bahasa Inggris dan bahasa Arab)



Coba: <https://w.wiki/AJ5u>

Jelajahi cakupan data leksikografis dengan lebih rinci



The screenshot shows a web browser window with the URL ordia.toolforge.org/lexical-category/. The page title is "Lexical categories". Below the title, there is a search bar with the text "Search after lemma or for" and a dropdown menu showing "Ordia". To the right of the search bar are links for "Languages", "Aspects", "Tools", and "Statistics".

Below the search bar, the page displays "Lexical categories" and a "Show 10 entries" dropdown menu. A table lists the following data:

Count	Lexical category
868548	noun
160700	adjective
146525	verb
49094	adverb
42718	proper noun
3757	postpositive adjective
3674	numeral
3118	root
2575	nominal locution
2546	interjection

Below the table, there is a link "Edit on query.Wikidata.org". At the bottom of the page, it says "Showing 1 to 10 of 289 entries" and a "Previous" button.

Ordia:
<https://ordia.toolforge.org/>

Jelajahi cakupan data leksikografis dengan lebih rinci

- Wikidata:Lexicographical coverage:
[d:Wikidata:Lexicographical_coverage](https://www.wikidata.org/wiki/d:Wikidata:Lexicographical_coverage)
- Wikidata:Lexicographical data/Statistics:
[d:Wikidata:Lexicographical_data/Statistics](https://www.wikidata.org/wiki/d:Wikidata:Lexicographical_data/Statistics)
- Wikimedia Grafana:
<https://grafana.wikimedia.org/goto/INWiGyyIq?orgId=1>

Untuk dapat membuat aplikasi yang bermanfaat, kita membutuhkan lebih banyak data yang lengkap dan beragam, serta lebih banyak orang untuk memeliharanya!



Pranala penting

Pranala penting

- Ikhtisar data leksikografis: [Wikidata:Lexicographical_data](#)
 - Mencoba peralatan yang ada
[Wikidata:Tools/Lexicographical_data](#)
 - Menyarankan ide untuk peralatan
[Wikidata:Lexicographical_data/Ideas_of_tools](#)
 - Ide dan contoh kueri
[Wikidata:Lexicographical_data/Ideas_of_queries](#)
- Membuat dan menyunting Leksem [Special:NewLexeme](#)
- Menyarankan atau mendiskusikan properti baru
[Wikidata:Property_proposal/Lexemes](#)
- Berdiskusi dengan komunitas mengenai cara memodelkan entri Leksem
 - Wiki talk page: [Wikidata_talk:Lexicographical_data](#)
 - Telegram: <https://bit.ly/4e8quK8>



Terima kasih banyak atas perhatian Anda!

Hubungi saya:

Raisha Abdillah

raisha.abdillah@wscproject.org

Atribusi

Presentasi ini dirilis dengan lisensi CC BY-SA 4.0.