

ウィキデータ語彙素

~デジタル公共財としての辞書のLOD作成プロジェクト

2026.3.15 LODチャレンジ2025授賞式



オープン・ナレッジ・ジャパン
東 修作/higa4

目次

本日も伝えたいこと

参考資料

本日お伝えしたいこと

言葉は自由ですか？→辞書のLOD作成へ

- ・言葉はその国の文化の礎であり、言葉の意味を定義したり文法的な分類を与えたりしたものが辞書である。
- ・辞書は時代とともに変化する言葉を写し出すとともに、言葉の意味を調べる際に参照されることで移ろいやすい言葉の意味を一定の範囲にとどめる基準線ともなっている。
- ・しかしながらその作成や更新には多大な労力が必要であり、その労力を回収できるビジネスモデルとして辞書の出版や関連事業が成立している。
- ・その結果、有償あるいは広告閲覧と引き換えに言葉の情報を入手しているケースが多い。
- ・辞書出版社や編纂者の努力や正当な権利を否定するものではないが、少なくとも誰でも自由に使える辞書の元になるデータはあってもよいのではないか。そのひとつが市民参加型で作るウィキデータ語彙素である。
- ・[ウィクショナリー](#)もよく似たプロジェクトであるが、ライセンス互換性の問題や(CC BY-SAライセンスであり、広く使われるためにはCC0がベターと考えられた)、データとして構造化されていないことなどから、2018年頃よりCC0の構造化された辞書データ、すなわちLODを作成するプロジェクトとしてウィキデータ語彙素が始まった。
- ・2026年にはウィキデータ語彙素等を自動翻訳に利用する抽象ウィキペディアの初期バージョンが姿を表す予定である。
- ・2025年9月にはウィキデータ全体が**デジタル公共財**として国連に認定され、日本でもデジタル庁資料によればデジタル公共財は「共助」領域のツールと位置づけられている。

ウィキデータの項目と語彙素

ウィキデータの要素には「項目(item)」と「語彙素(lexeme)」があります

要素	項目	語彙素
識別子	Q番号	L番号
登録対象	物事や概念	語彙素(語句、単語)
説明	百科事典の記事の単位で事実情報を構造化されたデータとして登録したもの	辞書の見出し語に相当する単位で語句を構造化されたデータとして登録したもの
実例	Q320943 :松戸市	L8218 :水

語彙素とは

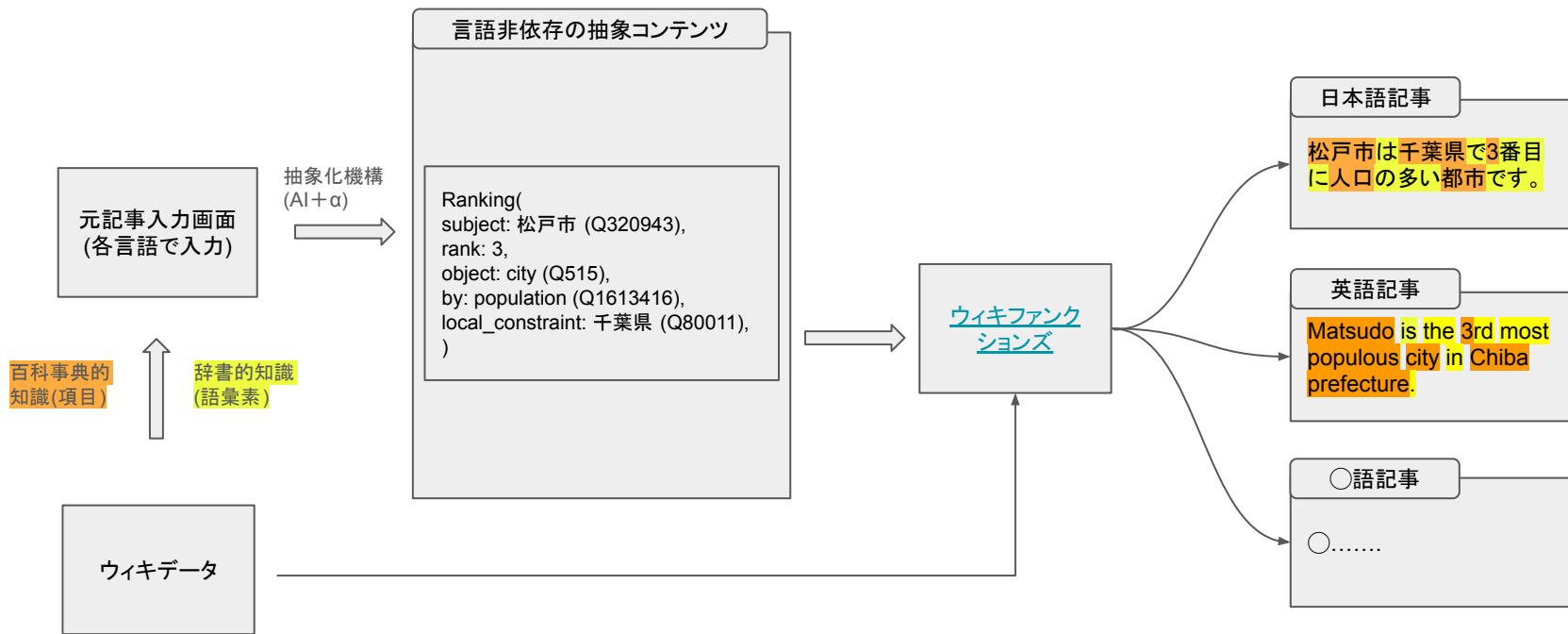
語、句、接頭辞といった、ある言語の語彙要素。辞書の見出し語に相当。

<例>

- ・品詞(名詞、動詞、形容詞、、、)
- ・慣用句
- ・熟語
- ・ことわざ
- ・接頭辞
- ・等々

抽象ウィキペディア (2026年登場予定) とウィキデータの関係

ー 抽象ウィキペディアの全体像はたぶんこんな感じ(2025/12/6時点の推測)



2026/2/20 抽象版ウィキペディアのプレビュー 1/4

Abstract Wikipedia:Q1348



 [Add languages](#) ▼

[Abstract Wikipedia](#) [Discussion](#) English ▼

[Read](#) [Edit source](#) [View history](#)  [Tools](#) ▼

Abstract content object

lead paragraph (Q8776414)

- >  [string to HTML fragment \(string of monolingual text \(defining role sentence \(→ wikidata item reference, capital city, West Bengal, language\)\)\)](#)
- >  [string to HTML fragment \(string of monolingual text \(Ordinal class location fragment \(→ wikidata item reference, 7, populous, city, India, language\)\)\)](#)

Auto-article preview in English

English

Kolkata

Kolkata is the capital of West Bengal. Kolkata is the seventh most populous city in India.

2026/2/20 抽象版ウィキペディアのプレビュー 2/4

Abstract Wikipedia:Q7679

[Add languages](#) ▾

[Abstract Wikipedia](#) [Discussion](#) [English](#) ▾

[Read](#) [Edit source](#) [View history](#) [★](#) [Tools](#) ▾

Abstract content object

lead paragraph (Q8776414)

- > [HTML raw content to HTML fragment \(intro for year \(1769 AD, language\)\)](#)
- > [section title \(header from item in plural \(event, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(May 9, language\), " – ", get label of item according to language fallbacks \(French conquest of Corsica, language\)\)\)\)](#)
- > [section title \(header from item in plural \(birth, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(August 15, language\), " – ", get label of item according to language fallbacks \(Napoleon, language\)\)\)\)](#)
- > [HTML raw content to HTML fragment \("<p/>"\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(September 14, language\), " – ", get label of item according to language fallbacks \(Alexander von Humboldt, language\)\)\)\)](#)
- > [section title \(header from item in plural \(death, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(February 2, language\), " – ", get label of item according to language fallbacks \(Clement XIII, language\)\)\)\)](#)

Auto-article preview in English

English

1769

1769 (MDCCLXIX) was a common year starting on Sunday of the Gregorian calendar, the 1769th year of the Common Era (CE) and Anno Domini (AD) designations, the 769th year of the 2nd millennium, the 69th year of the 18th century, and the 0th year of the 1760s decade.

Events

May 9 — French conquest of Corsica

Births

August 15 — Napoleon

September 14 — Alexander von Humboldt

Deaths

February 2 — Clement XIII

2026/2/20 抽象版ウィキペディアのプレビュー 3/4

Abstract Wikipedia:Q7679

[Add languages](#)

[Abstract Wikipedia](#) [Discussion](#) [English](#)

[Read](#) [Edit source](#) [View history](#) [Tools](#)

Abstract content object

lead paragraph (Q8776414)

- > [HTML raw content to HTML fragment \(Intro for year \(1769 AD, language\)\)](#)
- > [section title \(header from item in plural \(event, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(May 9, language\), " – ", get label of item according to language fallbacks \(French conquest of Corsica, language\)\)\)\)](#)
- > [section title \(header from item in plural \(birth, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(August 15, language\), " – ", get label of item according to language fallbacks \(Napoleon, language\)\)\)\)](#)
- > [HTML raw content to HTML fragment \("<p>"\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(September 14, language\), " – ", get label of item according to language fallbacks \(Alexander von Humboldt, language\)\)\)\)](#)
- > [section title \(header from item in plural \(death, language\)\)](#)
- > [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(February 2, language\), " – ", get label of item according to language fallbacks \(Clement XIII, language\)\)\)\)](#)

Auto-article preview in Dagbani

Dagbani

1769

1769 (MDCCLXIX) nyela tobu bali shee boni Alahari Gregorian kalendaafe 1769th zarj tinɗu'u ke: afe marɗan, afe 769th zarj tinɗu'u anɗuluni 2nd, afe 69th zarj tinɗu'u sekuli 18th, ke: afe 10th zarj tinɗu'u nyela puuni 1760.

Unable to render fragment

silimingoli May dabaawoi dali — French conquest of Corsica

N-dɔyi

silimingoli August biɛɣu pinaanu dali — Napoléon Bonaparte

silimingoli September biɛɣu pinaanahi dali — Alexander von Humboldt

Bahi

silimingoli February dabaayi dali — Clement XIII

「1769年」に関するダグバニ語の記事。ダグバニ語の単語Events」が利用できないため、エラーメッセージが表示されていますCC BY-SA 4.0

2026/2/20 抽象版ウィキペディアのプレビュー 4/4

Abstract Wikipedia:Q7679

🗺️ [Add languages](#) ▾

[Abstract Wikipedia](#) [Discussion](#) [English](#) ▾

[Read](#) [Edit source](#) [View history](#) ★ [Tools](#) ▾

Abstract content object

lead paragraph (Q8776414)

- > 🔗 [HTML raw content to HTML fragment \(intro for year \(1769 AD, language\)\)](#)
- > 🔗 [section title \(header from item in plural \(event, language\)\)](#)
- > 🔗 [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(May 9, language\), " – ", get label of item according to language fallbacks \(French conquest of Corsica, language\)\)\)\)](#)
- > 🔗 [section title \(header from item in plural \(birth, language\)\)](#)
- > 🔗 [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(August 15, language\), " – ", get label of item according to language fallbacks \(Napoleon, language\)\)\)\)](#)
- > 🔗 [HTML raw content to HTML fragment \("<p>"\)](#)
- > 🔗 [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(September 14, language\), " – ", get label of item according to language fallbacks \(Alexander von Humboldt, language\)\)\)\)](#)
- > 🔗 [section title \(header from item in plural \(death, language\)\)](#)
- > 🔗 [string to HTML fragment \(concatenate many strings \(Z881 \(String, display Day of Roman Year \(February 2, language\), " – ", get label of item according to language fallbacks \(Clement XIII, language\)\)\)\)](#)

Auto-article preview in German

1769

1769 (MDCCLXIX) war ein Gemeinjahr das an einem Sonntag begann. Es war ein Jahr im Gregorianischem Kalender. Es war das 1769. Jahr unserer Zeitrechnung und nach Christus, das 769. Jahr des 2. Jahrtausends, das 69. Jahr des 18. Jahrhunderts, und das 10. Jahr der 1760er Jahre.

Veranstaltungen

9. Mai – French conquest of Corsica

Geburten

15. August – Napoleon Bonaparte

14. September – Alexander von Humboldt

Tode

2. Februar – Clemens XIII.

ウィキペディア記事における語彙素の網羅状況サンプル

白血病 (はっけつびょう、**英語**: Leukemia)は、**遺伝子変異**を起こした**造血細胞** (白血球系細胞)が**骨髄**で無限に増殖して正常な造血を阻害し、多くは骨髄のみにとどまらず**血液**中にも白血病細胞があふれ出てくる**血液疾患**である。別名「**血液のがん**」ともいわれる^{[1][2]}。

白血病細胞が造血の場である**骨髄**を占拠するために**造血**が阻害されて正常な**血液細胞**が減るため**感染症**や**貧血**、**出血**症状などの症状が出やすくなり、あるいは**骨髄**から血液中にあふれ出た白血病細胞がさまざまな**臓器**に**浸潤**(侵入)して障害することもある。

治療は、**抗がん剤**を中心とした**化学療法**と**輸血**や感染症対策などの**支持療法**に加え、難治例では**骨髄移植**や**臍帯血移植**などの**造血幹細胞移植**治療も行われる。

白血病は、**急性骨髄性白血病** (AML)、**急性リンパ性白血病** (ALL)、**慢性骨髄性白血病** (CML)、**慢性リンパ性白血病** (CLL)の4つに分けられる。

ウィキペディアトップページから
適当に選んだ記事の冒頭

ウィキペディア記事における語彙素の網羅状況サンプル

白血病(はっけつびょう、英語: Leukemia)は、遺伝子変異を起こした造血細胞 / 白血球系細胞 / が 骨髄で無限に増殖して正常な造血を阻害し、多くは骨髄のみにとどまらず血液 / 中 / にも 白血病細胞があふれ出 / て / くる 血液 / 疾患である。別名 / 血液のがん / とも / い / わ / れ / る / 1 / 1 / 2 / 1。

白血病細胞が造血の場である骨髄を占拠するために造血が阻害されて正常な血液細胞が減るため 感染症 / や / 貧血 / 出血 / 症状 / など / の / 症状が出 / やす / く / な / り、ある / い / は 骨髄から血液 / 中 / に あ / ふ / れ / 出 / た 白血病細胞がさま / ざま / な / 臓器 / に 浸 / 潤 / し / て 障 / 害 / す / る / こ / と / も / あ / る。

治療は、抗がん剤を中心とした化学療法と輸血 / や / 感染症 / 対策 / など / の / 支持療法 / に / 加 / え、難 / 治 / 例 / で / は 骨髄移植 / や / 臍帯血 / 移植 / など / の 造血幹細胞移植 / 治療 / も / 行 / わ / れ / る。

白血病は、急性骨髄性白血病 / (/ AML /) / 急性リンパ性白血病 / (/ ALL /) / 慢性骨髄性白血病 / (/ CML /) / 慢性リンパ性白血病 / (/ CLL /) / の / 4 / つ / に / 分 / け / ら / れ / る。

- : ウィキデータ項目あり
- : ウィキデータ語彙素あり
- : いずれも未登録

語彙素は形態素(言葉として意味を持つ最小の単位)レベルで登録されている

The Surrounding Ocean(語彙素構造の可視化ツール)

乳製品 / にゅうせいひん
 — 日本語, 名詞 — 家畜の乳から作られる製品。

チンする / ちんする — 日本語, 動詞 — 電子レンジで温める。

チーズケーキ — 日本語, 名詞
 ブルーチーズ — 日本語, 名詞

formaggio — イタリア語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品

チーズ / 乾酪 — 日本語, 名詞 — 乳製品のひとつ。



 / **nIm wIb ngogh** — クリントン語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品



cheese — 英語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品

チーズケーキ — 日本語, 名詞 — チームがかかったケーキ。

ブルーチーズ — 日本語, 名詞 — アオカビのついたチーズ。



The Surrounding Ocean(語彙素構造の可視化ツール)

乳製品 / にゅうせいひん
— 日本語, 名詞 — 家畜の乳から作られる食品。

チンする / ちんする — 日本語, 動詞 — 電子レンジで温める。

チーズケーキ — 日本語, 名詞

ブルーチーズ — 日本語, 名詞

formaggio — イタリア語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品。

チーズ / 乾酪 — 日本語, 名詞 — 乳製品のひとつ

ngogh — クリントン語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品

cheese — 英語, 名詞 — チーズ; 乳を原料とし、凝固や発酵などの加工をしてつくられる食品。

チーズケーキ — 日本語, 名詞 — チーズがかかったケーキ。

ブルーチーズ — 日本語, 名詞 — チーズの一種

文 A ? 🔍 *

TheSurroundingOcean(語彙素構造の可視化ツール)

(L683699) **チーズ** | 乾酪 ✎編集

ja | ja-hani

言語: 日本語
語彙範疇: 名詞

文

分類 ✎ 0件の情報源

以下に由来 ✎ cheese ✎ 1件の情報源

識別子

デジタル大辞典ID (goo辞書) ✎ 140904 ✎ 0件の情報源

表記名 チーズ

語義

L683699-S1

英語 ✎ yellow or white, creamy or solid food made from the pressed curds of milk

日本語 ✎ 乳製品のひとつ。

L683699-S1の文

画像 ✎ 0件の情報源



Four cheese plates on the haven.jpg
5,184 × 3,888 ; 7.22メガバイト

この語彙に対応する項目 ✎ チーズ ✎ 0件の情報源

上位語 ✎ 乳製品/にゅうせいひん (日本語) - 家畜の乳から作られる製品。 ✎ 0件の情報源

翻訳

アッカド語: eqidu, gubnatu	スウェーデン語: ost	ヘブライ語: חֶמְסָה/חֶמְסָה
アルバニア語: djath/djathë	スペイン語: queso	ポーランド語: ser
イタリア語: formaggio	ダバニ語: waagashi, waagashi	ポルトガル語: queijo
イボ語: chiiz	チェコ語: sýr	マリシール語=バサマコティ語: ocis
ウクライナ語: сир	デンマーク語: ost	マレー語: keju/چوکی
エストニア語: juust	ドイツ語: Käse	モンゴル語: бяснар/бяснар
エスペラント語: fromaĝo/fromaĝo/fromaĝo	ハウザ語: cukwi/كوكي	ラテン語: caseus
オランダ語: kaas	バスク語: gazta	ロシア語: сыр, сырок
ガリシア語: queixo	ハンガリー語: sajt	英語: cheese
クリンゴン語: nim wib ngogh/□□□□□□□□	パンジャブ語: ਸੌਰ/ਸੌਰ	客家語: foh-maa-se/福马瑟/福馬瑟 (モーリシャス)
シヌメール語: ga'ar/گائار, ellangunu/اللانغونو, birgun/بيرگون	フィンランド語: juusto	広東語: 芝士/芝士
	ブークモール: ost	西フリジア語: tsjis
	フランス語: fromage	中央ドゥスン語: keju
	ブルトン語: keuz	

ブラシル手話:



語形

L683699-F1 | **チーズ** | 乾酪 ✎

ja | ja-hani

文法的特徴

L683699-F1の文

修正ヘボン式ローマ字表記 ✎ chizu ✎ 0件の情報源

日本語語彙素の編集回数

アカウント	編集回数
Higa4	161,025
*	29,547
*	4,517
*	4,178
*	3,903
*	3,777
*	2,864
*	1,866
*	1,733
*	1,013

統計

言語別語彙素数(2026/3/14現在)

No.	言語	語彙素数	No.	言語	語彙素数
1	German	241,890	11	Arabic	53,714
2	Russian	102,240	12	Modern Greek	43,333
3	Danish	98,497	13	Bokmål	40,694
4	Estonian	83,223	14	Aragonese	38,665
5	English	73,510	15	Czech	30,990
6	Malayalam	67,367	16	Hebrew	29,636
7	Italian	64,882	17	Nynorsk	26,039
8	Spanish	63,954	18	Japanese	24,141
9	Latin	56,329	19	Basque	22,949
10	Swedish	54,068	20	French	20,930

参加者募集！

- ・ウィキデータ語彙素における日本語の登録数は2026年3月現在24,000件程度。小型国語辞典の収録語数は7万件程度とされていることから、そのあたりがひとつの目安。
- ・現時点ではアクティブな編集者数が限られておりまだ実用レベルには質・量とも不十分であるため、その目的や現状をお伝えすることで日本語語彙素の充実に協力して頂ける参加者が増えることを期待。

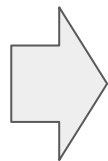
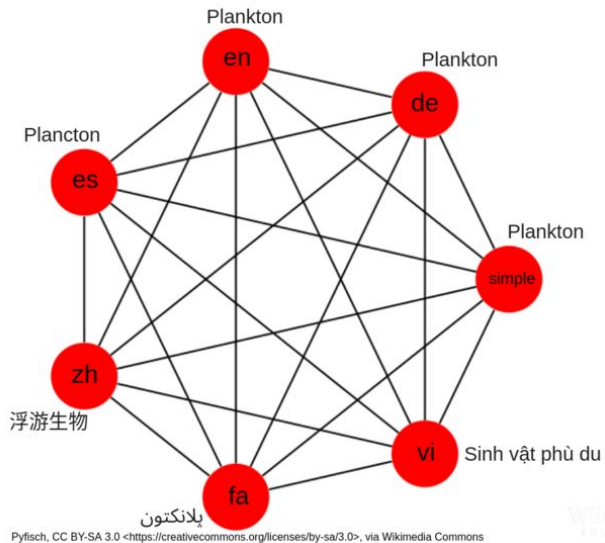
<やることの例>

- ・ウィキデータ語彙素の登録、改善
- ・ウィキデータ項目の構造整理
- ・ウィキファンクションズの日本語用関数の開発
例: 西暦和暦変換, 与えられた動詞の活用形を返す
- ・日本語の語彙素データ活用/編集アプリ開発
例: [Dagbanli Dictionary](#), [Scribe-iOS](#)
- ・などなど

參考資料

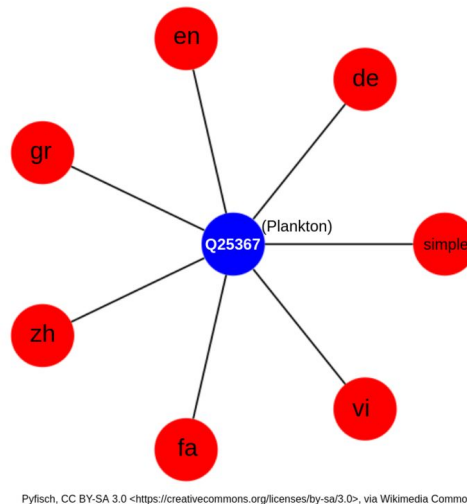
ウィキペディアとウィキデータの関係 (はじまり)

Interlanguage links (aka interwiki links)



ウィキデータ
以後

Sitelinks



Wikipedia (92 entries)

af	Plankton
als	Plankton
ar	عوالق
ast	Plankton
az	Plankton
be_x_old	Плянктон
be	Плянктон
bg	Плянктон
bn	প্লান্কটন
br	Plankton
bs	Plankton
ca	Plankton
cs	Plankton
cy	Plankton
da	Plankton
de	Plankton
el	Πλανκτόν
en	Plankton
eo	Planktono
es	Plankton
et	Plankton
eu	Plankton
fa	پلانکتون
fi	Plankton
frr	Plankton
fr	Plankton
ga	Planktón
gd	Meanbh-bheò
gl	Plancto

言語間リンクは異なる言語のウィキメディアプロジェクトにおいて手動で記事を他の言語版に接続する。

サイトリンクは他言語および他のウィキメディアプロジェクトにあるすべての関連記事にリンクするためにウィキデータの項目をハブとして利用する。

抽象ウィキペディアの目的と位置づけ

- ・言語版により記事数の差が大きいことから、そのギャップを翻訳記事で埋めて広く知識の共有を図ろうとするもの。
- ・各言語版間のギャップを埋める手動翻訳は膨大な作業であり、さらに翻訳後も改訂時の同期を取り続けるのは至難の業。そこで抽象的な構造を持つコンテンツをひとつ作ってそこから各言語記事に自動翻訳して表示しよう、というもの。
- ・既存の記事やその作成/手動翻訳方法を置き換えるものではなく、あくまで将来新しい記事を作成する場合に自動的に各言語に翻訳される抽象コンテンツ「も」作成できるようにするもの。

ウィキデータ語彙素のはじまり

- ・世界中の語彙をCC0の構造化データとして集積するプロジェクトとして2018年に開始。
- ・主にライセンスの問題よりウィクショナリーとは別にフルスクラッチでデータ整備が始まった。

語彙素の構造

見出し

(L1529535)	災害	さいがい	
	ja	ja-hira	
言語: 日本語 語彙範疇: 名詞			
文			
+ 追加			

語彙素全体に係る属性


注釈欄ではできるだけ語彙の定義を書く。複数の語義がある場合は気づいた時に都度追加すれば良い。定義が難しい場合は別の単語による言い換えでもひとまずは良いが、最終的には定義を書いて、言い換えの単語は別の見出し語として登録した上で、同義語プロパティで関連付ける形が良い。

意味

語義			
L1529535-S1	日本語	天災や人災による被害	
L1529535-S1の文			
+ 追加			
+ 語義を追加			

語義に係る属性

変化形

語形			
L1529535-F1	災害	さいがい	
	ja	ja-hira	
文法的特徴			
L1529535-F1の文			
+ 追加			
+ 語形を追加			

動詞や形容詞はその活用形をすべて登録。名詞は表記揺れ、旧字体、訓読み / 音読み、などの表記のバリエーションを登録。

語形に係る属性
名詞: 読み、連濁等
動詞: 活用形等

日本語語彙素の登録一覧

日本語の見出し語一覧(見出し語順)

```
SELECT ?lexemeld ?lemma WHERE {  
  ?lexemeld dct:language wd:Q5287;  
  wikibase:lemma ?lemma.  
}  
ORDER BY (?lemma)
```

上記実行結果は2026/3/14時点で46,932件。
ひとつの語彙素に対して、見出しは複数(ほぼ 2倍)登録されている。

統計

比較対象	総数	時点	情報源
ウィキペディア(日本語版・記事)	1,481,844	2025/11/29	総記事数
ウィキデータ(全体・項目)	119,676,762	2025/11/29	コンテンツページ数
ウィキデータ(全体・語彙素)	1,508,230	2025/11/29	言語別語彙素数
ウィキデータ(日本語・語彙素)	21,463	2025/11/29	同上
ウィキデータ(日本語・語彙素見出し語) ^注	41,755	2025/11/29	日本語の見出し語一覧
ウィクショナリー(日本語版・項目)	461,830	2025/11/29	コンテンツページ数
ウィクショナリー(日本語版・日本語語句)	127,819	2025/11/29	カテゴリ「日本語」にあるページ
小型国語辞典(紙媒体・収録語数)	7万～10万	2024/7	国語辞典の種類と選び方

注: 語彙素は通常複数の見出し語を持つ(例:「春」と「はる」)

統計

言語別語義数(2025/11/29現在)

No.	言語	語義数	No.	言語	語義数	No.	言語	語義数
1	English	55,092	11	German	19,469	21	French	11,803
2	Bokmål	49,825	12	Spanish	18,985	22	Northern Sami	10,045
3	Aragonese	33,758	13	Egyptian	18,851	23	Fula	9,062
4	Nynorsk	31,957	14	Russian	17,988	24	New Persian	8,981
5	Swedish	30,867	15	Sumerian	16,947	25	Punjabi	8,313
6	Basque	30,758	16	Danish	16,913	26	Albanian	8,145
7	Japanese	25,796	17	Akkadian	14,970	27	Hebrew	7,095
8	Italian	25,471	18	Dagbani	14,186	28	Igbo	5,691
9	Czech	24,269	19	Bangla	13,995	29	Portuguese	5,016
10	Persian	19,624	20	Malayalam	12,544	30	Estonian	4,777

統計

言語別語形数(2025/11/29現在)

No.	言語	語形数	No.	言語	語形数	No.	言語	語形数
1	Estonian	2,802,122	11	Italian	522,922	21	Nynorsk	131,524
2	German	2,020,920	12	Ukrainian	508,127	22	Bangla	87,104
3	Arabic	1,820,983	13	Aragonese	443,566	23	Northern Sami	74,063
4	Basque	1,257,084	14	Hebrew	437,695	24	Modern Greek	70,993
5	Russian	1,247,008	15	Indonesian	413,309	25	Portuguese	48,356
6	Latin	1,199,242	16	Swedish	355,064	26	Japanese	39,287
7	Czech	875,111	17	French	335,078	27	Persian	39,253
8	Malayalam	753,106	18	Bokmål	240,987	28	Polish	32,220
9	Spanish	694,931	19	Slovak	235,263	29	Punjabi	22,696
10	Danish	652,014	20	English	148,142	30	Dutch	21,555

統計

言語別文例数(2025/11/29現在)

No.	label	文例数	No.	label	文例数	No.	label	文例数
1	Swedish	5,605	11	German	925	21	Kadazandusun	288
2	Igbo	4,870	12	Twi	726	22	Ewe	284
3	Japanese	3,597	13	Egyptian	594	23	Gun	276
4	Polish	2,045	14	Portuguese	550	24	New Persian	256
5	Danish	2,024	15	Indonesian	537	25	Fula	204
6	Banjar	1,840	16	Russian	444	26	Spanish	179
7	Dagbani	1,494	17	Southern Min	392	27	Turkish	177
8	Lombard	1,248	18	Italian	391	28	Tyap	170
9	Standard Chinese	1,126	19	Czech	383	29	Hausa	169
10	English	1,061	20	Bangla	338	30	Baoulé	159

統計

言語別画像数(2025/11/29現在)

No.	言語	画像数	No.	言語	画像数	No.	言語	画像数
1	Russian	5,321	11	German	613	21	Gun	214
2	Danish	2,744	12	Dagaare	526	22	Southern Min	210
3	Spanish	2,182	13	English	460	23	Banjar	210
4	Dagbani	1,759	14	Punjabi	329	24	Indonesian	207
5	Japanese	1,413	15	Malayalam	323	25	Tyap	201
6	Bokmål	1,031	16	Central Bikol	323	26	Malay	172
7	Polish	1,021	17	Fula	301	27	Czech	171
8	Nynorsk	907	18	Tamil	238	28	Baoulé	156
9	Igbo	671	19	Ewe	230	29	Kadazandusun	131
10	Twi	636	20	French	224	30	Portuguese	122

統計

言語別音声ファイル数(2025/11/29現在)

No.	言語	音声ファイル	No.	言語	音声ファイル	No.	言語	音声ファイル
1	German	86,331	11	Dagbani	1,631	21	Banjar	432
2	Bangla	36,371	12	Persian	1,459	22	Bavarian	293
3	Dutch	18,112	13	Basque	1,259	23	Hindustani	289
4	English	11,959	14	Quechua	1,174	24	Japanese	233
5	Ukrainian	11,345	15	Tamil	957	25	Danish	231
6	Russian	9,063	16	New Persian	890	26	Gun	210
7	Malayalam	8,408	17	Spanish	621	27	Toki Pona	192
8	Polish	7,809	18	Esperanto	572	28	Punjabi	183
9	French	4,847	19	Swahili	558	29	Arabic	180
10	Swedish	3,750	20	Czech	436	30	Turkish	175

統計

日本語の語彙範疇(≡品詞+α)別語彙素数(2025/11/29現在)

No.	語彙範疇	件数	No.	語彙範疇	件数	No.	語彙範疇	件数
1	名詞	16,426	11	助詞	77	21	助動詞	28
2	動詞	2,347	12	数詞	72	22	数量詞	13
3	副詞	555	13	助数詞	71	23	人称代名詞	8
4	固有名詞	542	14	接頭辞	67	24	接辞	6
5	形容詞	298	15	ことわざ	65	25	不変化詞	4
6	慣用句	254	16	句	57	26	文字	4
7	形容動詞	127	17	代名詞	49	27	音節	3
8	接尾辞	99	18	感動詞	39	28	形態素	2
9	漢熟語	98	19	連体詞	37	29	字母	1
10	接続詞	80	20	連語	29	30	助字	1

登録手順のさわりだけ

まずその単語が登録済みかどうかを確認してください

・[ウィキデータ画面](#) 右上の検索窓で「L:単語」のように検索します。Lは大文字小文字を問いません。

<登録済みの場合>

<未登録の場合>

出てきた候補から選択します

「新しい語彙素を作成」から新規作成します



特別ページ

検索結果

ウィキデータの項目を各サイトでのページ名から検索するためには、Special:ItemByTitleを使ってください。

Q L:歩く

高度な検索:

検索範囲:

歩く/あるく (L15111)
日本語, 動詞…
文: 7, 節形: 7 - 2025-09-11T12:15:17

思い歩く/おもいあるく (L1080603)
日本語, 動詞…
文: 5, 節形: 1 - 2024-05-05T06:31:15

てくてく歩く/てくてくあるく (L1191912)
日本語, 慣用句…
文: 2, 節形: 0 - 2025-08-30T14:09:26



特別ページ

検索結果

ウィキデータの項目を各サイトでのページ名から検索するためには、Special:ItemByTitleを使ってください。

Q L:浸水

高度な検索:

検索範囲:

問い合わせに合致する検索結果はありませんでした。「浸水」に対して新しい語彙素を作成することもできます。

なぜ語彙素か～モチベーション

・楽しい

- ・言葉の意味や構造を解きほぐして立体的にデータ化する作業が楽しい

『[笑う](#)』の意味もよく考えるとひとつではない。

→「おかしくて笑う」「見下して笑う」「膝などがガクガクする」

『[階段](#)』とは何？

→例:「高さの異なる2点間を歩いて行き来できるように設けられた段状の設備」

『[春](#)』という言葉の上位概念は？

→「四季」「季節」

・誰かの役に立ちそう

- ・[抽象ウィキペディア](#)

- ・生成AIの学習データ/RAG

- ・電子辞書など作りたい人

参加方法

いろいろなやり方で参加できます

1 語彙素の新規追加

1.1 保護期間満了の辞書などから転記

- ・まずは国語辞典として現代でも使われている単語を選び、定義の意味、字体、仮名遣いなどは現代のものに変換
- ・あらゆる語彙を受け入れる形になっており、古語、方言、ことわざ等も登録できる

1.2 自分の知識で追加

- ・ウィキペディア記事や書籍などから単語を拾い、未登録であれば定義などを自分で考えながら登録する
- ・外来語は比較的未登録が多い
- ・慣用句も結構重要(例えば「[手を加える](#)」が「add hands」と訳されないように)

参加方法

2 意味的な整備

2.1 同義語(類義語)、対義語を充実

2.2 「文例(用例)」を充実(自分の知識でも可。青空文庫等の引用がベター。)

2.3 語義に対応する「画像」をcommonsより追加

2.4 「この語義に対応する項目」を追加

2.5 「複合語の成分」を追加

2.6 翻訳(相互に)

参加方法

3 機械的な整備

3.1 活用語形を記載

- ・名詞: 見出しと同内容を転記し、さらに表記ゆれ、異体字などがあれば追加
- ・動詞、形容詞: 各活用形を記載 ([ツール](#)あり)

3.2 別の表現を記載

- ・ローマ字表記
- ・IPA発音、点字
- ・音声ファイル

語彙素の課題

1 全般的に

- ウィキデータ項目と紐づいた知識構造(ナレッジグラフ)の部分がまだまだ整備不足。語彙数を増やすだけでなく、概念の関係性の整備が必要。
- 言語ごとに異なる語彙素モデルは各国それぞれで試行錯誤中。最大のユースケースである抽象ウィキペディアがプロトでも動くようになればあるべき姿が見えてくるはず。
- 語義は言語間のマッピング(翻訳)ができる粒度にして行く必要がありそう。

語彙素の課題

2 日本語に関して

○量と質の両面での充実

- ・ご興味ありましたらぜひどうぞ

○紙の辞書とは異なるデータならではの悩みどころ

- ・同一の語彙素内で扱うか、別の語彙素としてエントリーを分けるのかの判断が悩ましいものの整理
 - ・名詞と複合動詞(「忬度」と「忬度する」)→別が良さそう？
 - ・名詞と形容動詞(形容動詞を品詞と見るのか「名詞」+「助詞」とみるのか)
 - ・受動態(「投げる」と「投げられる」)→前者は「動詞」、後者は「連語」とするのが良さそう？
 - ・敬語表現→受動態と同じ？
 - ・丁寧表現(「だ」と「です」「ます」)
 - ・同じ漢字で意味が異なるもの(「一味(仲間)」と「一味(ひとつの味)」)→別が良さそう？

○抽象ウィキペディアに向けて

- ・[日本語を操作する関数](#)(ウィキファンクションズ)の開発→まだ片手で数えるほどしか無い

○著作権保護の範囲や期間満了の判断

抽象ウィキペディアへの期待と不安

- ・言語に依存しない抽象コンテンツの記述中に埋め込まれる事実情報が、ウィキデータの項目や語彙素である。
- ・抽象コンテンツは生成AIのような確率論的なロジックではなく、入力が決まれば出力も決まる関数的なロジック、すなわちウィキファンクションズで各国語で読める記事に翻訳される。
- ・すでに生成AIによる機械翻訳が(一見?) 实用レベルに達している現在、ひと昔前の手法と思える関数的な変換手法がどこまで自然な訳文を作り出せるのか、ウィキデータの質と量にはまだ課題があり、抽象ウィキペディア構想も手探りで進んでいる中、成功への鍵はまだ足りていない感触。[ベクターデータベース](#)とか？
- ・しかしながら不確実性を排除するために独自の道を切り開こうとしているプロジェクトへの参加は面白く、今後とも主にデータの面で参加を続けたい所存。

祝！デジタル公共財 (Digital Public Goods) の仲間入り



[Wikidata 2025/9](#) ←NEW!

[Wikipedia 2025/1/31](#)

[OpenStreetMap 2023/11/27](#)

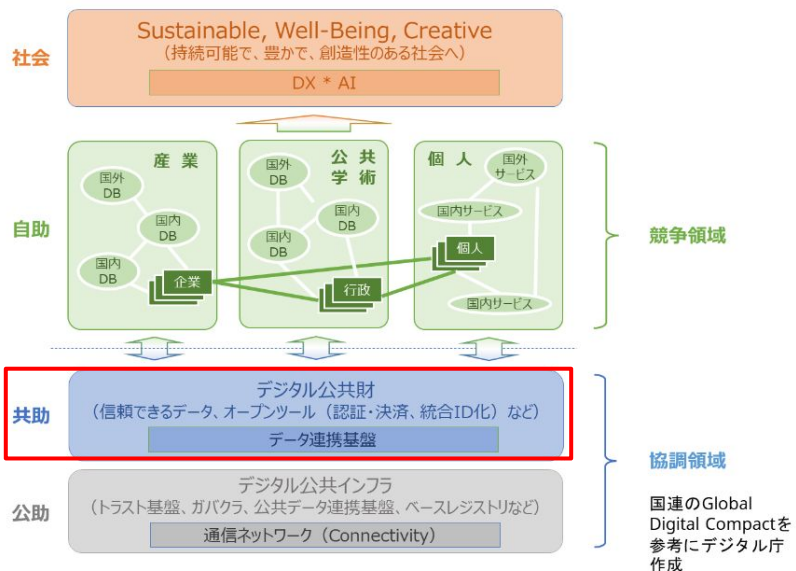
The UN-endorsed Digital Public Goods Alliance has added Wikipedia and Wikidata to its registry, affirming the important role their volunteer contributors play in making the internet better for everyone.

UPDATE (29 October 2025) – Last month, Wikidata became the second Wikimedia project to be officially recognized as a digital public good (DPG) by the Digital Public Goods Alliance (DPGA). Wikidata is a free and open knowledge base, which can be read and edited by both people and machines, and acts as central storage for the structured data (including text, images, and multimedia) of the other Wikimedia projects. Its structured data is also available for use under a free license to other external applications and services

社会の中のデジタル公共財~「共助」領域の位置づけ

国連Global Digital Compactをベースとした世界観

- デジタル公共インフラは、マイナンバーカードなどの認証基盤、ガバメントクラウドなど、地域のDXを進める上で不可欠となる公共インフラが相当します。他方、通念上の公共性が高いものであっても、ソフトウェアやデータの場合、簡単にコピーが出来るしまうので、なかなか公共財としての取扱が得られません（“限界費用がゼロであることに伴う自然独占が成立しやすい財”）。
- しかし現実には、広く使われるべきソフトウェアやデータが、結果的に購買力のある地域に導入が限定されてしまい、広く普及せず、結果として、地域の暮らしのDXのブレーキとなっているケースが多々見られます。このため、国連も、デジタル公共インフラとデジタル公共財を区別し、どのようにすればデジタル公共財の普及・活用を促すことが出来るのか、課題提起を行っているところです。



辞書と著作権(ウィキメディア財団による予備的な見解)

著作権保護対象	著作権保護対象外
マイクロ構造とマクロ構造(著者が配置を選択できないような、特定の言語に対するやり方が決まっている場合を除く)	見出し語
創造性の余地がある定義	専門用語や定型表現の定義
語用論的情報	文法情報
百科事典的な情報と例文	連語と定型表現

表内の単語:	意味:
マイクロ構造	マイクロ構造とは、各カテゴリ内の各エントリのレイアウトと配置です。たとえば、太字の単語を表示するように選択し、最初に発音ガイド、次に文法情報、次に定義、次に例文を使用するように選択するようなことです。
見出し語	定義の題目。通常は辞書の太字部分で、発音や註解を含みます。
マクロ構造	辞書の全体的な構造(アルファベット順の配置ではない)。たとえば、目次を挿入した後にエントリー一覧を挿入する構造などです。
連語	連語とは、一緒になって辞書で一緒に定義されなければならない単語のことです。例えば、「heavy」rainは「powerful」rainとは表現できません。
語用論的 情報	単語の意味に文脈を提供し、著者がどのように意味を伝達したいかを読者に示します。
定型表現	辞書に不可欠であり、すべての辞書に共通する用語と表現。たとえば、「All of a sudden (突然)」は、すべての辞書に共通する決まったやり方でのみ定義できます。

語彙素登録内容の元になるリソース(私見)

<利用できるのはおおまかには以下>

- ・著作権保護期間満了の辞書等
- ・保護期間内であっても登録する際に満たすことのできる条件(出典表記など)の下に二次利用が許諾されているもの
- ・保護期間内であっても著作性の無い部分(判断には注意)
- ・著作性が認められる可能性がある内容を参考に自分なりの表現に置き換えたもの
- ・自分の知識(他の、より信頼性の高いリソース等で随時上書きされ得る前提)

<思うに>

- ・サーバーの所在が公表されておらず著作権保護期間の判断は曖昧なところがある。
- ・保護期間や著作権有無に関わらず辞書に記載された言葉の定義は自分の言葉で置き換えれば侵害とされない可能性が高い。
- ・NDLデジコレなどで著作権保護期間満了と判断された辞書に記された言葉の定義はその当時の世相が反映されていたり、旧仮名遣いや旧字体が使われているなど、現代の辞書としてはそのまま使えないことが多いため、必然的にまるごとコピーではなく編集者の言葉で置き換えた内容になる。そのため、仮に日本より長期の保護期間が及ぶ場合であったとしても著作権侵害とされない表現の置き換え等が行われていれば問題ない。
- ・出典があるに越したことはないが、実際は現代の言葉をCC0で利用できるリソースは少ないため、出典なしに自分(母語としての話者)の知識を元に登録することも問題ない。

ウィクシヨナリーとウィキデータ語彙素 画面の比較

あこがれる

🌐 6 種の言語版 ▾

項目 [議論](#) [閲覧](#) [編集](#) [履歴を表示](#) ☆

日本語 [\[編集\]](#)

動詞 [\[編集\]](#)

あこがれる【[憶れる](#)、[憶れる](#)】

- (自動詞) 対象に向け、そわそわと落ち着かないくらい、心が引かれる。理想のものとしてあこがれる。
 - そして芸術は、この地上的なものが天上的なものを夢見、仮にわけもち、あこがれることで現象を遊離して象徴の中に入ることとなるのである。(中井正一『現代美学の危機と映画理論』)

活用 [\[編集\]](#)

活用と結合例 [鳴す]						
あこがれる 動詞活用表 (日本語の活用)						
ラ行下一段活用						
語幹	未然形	連用形	終止形	連体形	仮定形	命令形
あこが	れ	れ	れる	れる	られ	れろよ

各活用形の基礎的な結合例		
意味	語形	結合
否定	あこがれない	未然形 + ない
意志・勧誘	あこがれよう	未然形 + よう
丁寧	あこがれます	連用形 + ます
過去・完了・状態	あこがれた	連用形 + た
言い切り	あこがれる	終止形のみ
名詞化	あこがれること	連体形 + こと
仮定条件	あこがれば	仮定形 + ば
命令	あこがれろ あこがれよ	命令形のみ

語源 [\[編集\]](#)

- 古典日本語: [あくがる](#)より

関連語 [\[編集\]](#)

- 類義語: [憧憬](#)

翻訳 [\[編集\]](#)

- 英語: [admire](#) ^(en)

(L1219823)

憶れる

ja

あこがれる

ja-hira

[編集](#)

言語: 日本語
語彙範疇: 動詞

文

掲載している情報源 [編集](#) 👍

大日本国語辞典 [編集](#)

詳細情報URL <https://dl.ndl.go.jp/oid/254645/1/33>

0件の情報源 [+ 情報源を追加](#)

[+ 追加](#)

活用の種類 [編集](#) 👍

ラ行下一段活用

0件の情報源 [+ 情報源を追加](#)

[+ 追加](#)

以下の漢字を含む [編集](#) 👍

憶

0件の情報源 [+ 情報源を追加](#)

[+ 追加](#)

動詞の種類 [編集](#) 👍

自動詞

0件の情報源 [+ 情報源を追加](#)

[+ 追加](#)

語幹 [編集](#) 👍

憶 (日本語)

0件の情報源 [+ 情報源を追加](#)

[+ 追加](#)

ウィキデータ語彙素からウィクショナリーへの連携例

Haus 74 jezika

Stranica Razgovor Čitaj Uredi Vidi povijest ☆

Haus (njemački jezik) [uredi]

izgovor: IPA: [ˈhaʊs], poslušaj

prijevod:

imenica

(1.1) kuća



[1] Häuser in London



[1] Haus in Deutschland

	das Haus	
Padež	Jednina	Množina
Nominativ	das Haus	die Häuser
Genitiv	des Hauses	der Häuser
Dativ	dem Haus	den Häusern
Akuzativ	das Haus	die Häuser

ウィキファンクションズが自動生成した表

< 編集画面 >

```

== Haus ({{njemački jezik}}) ==
[[Datoteka:Abingdon House.jpg|mini|desno|250px|[1] 'Häuser' in [[London]]]]
[[Datoteka:Radebeul Einfamilienhaus Clara Toller (2).JPG|mini|desno|250px|[1] 'Haus' in [[Deutschland]]]]
{{izgovor}} {{IPA|ˈhaʊs}}, {{audio|De-Haus2.ogg|poslušaj}}
{{prijevod}}
{{imenica|jezik=de|rod=|vrsta=}}
: (1.1) [[kuća]]
{{#function:Z28602|L2957|}}
    
```

ウィキファンクションズの関数ID: Z28602)に引数として語彙素ID: L2957)を渡して呼び出している

Kategorije: [njemački \(indeks\)](#) | [Njemačke imenice](#)

語彙素用プロパティ

1)全体に係るもの

分類: 謙讓語

[掲載している情報源 \(P1343\)](#):

以下に由来 (P5191): 語の最も直接的な由来となる語彙素

— 派生様式 (P5886): 略語、アナグラム、

語根 (P5920): 接頭辞や接尾辞のない単語の部分

<動詞の場合>

活用の種類:

語幹 (P5187): 主題の語彙素の語幹

動詞の種類 (P9295) : 自動詞、他動詞

文例:

以下と同一視される (P11577): この語彙素は別の語彙素と同じもしくは綴りの変種だと考える情報源がある(簡易な字への置き換えなど?)例: 母、おっかあ

名詞形 (P7706)

語彙素用プロパティ

2) 語義に係るもの

動詞の行為名(P9970)例:「泳ぐ」に対する「水泳」

画像(P18)

この語義に対応する項目P5137)

基本動詞(P5975) 例:「嘲笑う」に対する「笑う」

詳述語義(P6719)説明的な語義で、基本動詞に付加された際に、より詳細な主題の語義を決めるもの(修飾子)

特性(P1552): 夏の季語

同義語 (P5973)

対義語 (P5974)

上位語(P6593) 例:「犬」(この語彙)に対する「動物」(別の語彙)

以下の付属語(P8471) 例:「slowly」(この語彙、副詞)に対する「slow」(元の関連する語彙、形容詞)

註解(P8394): 情報源から語義の注釈を引用できる場合

類別詞(P5978): この語義 (名詞) の助数詞 (手紙に対する「通」)

意味上の性(P10339): 男性、女性

語彙素用プロパティ

3)語形に係るもの

文法的特徴: 体言、畳語(諾諾)、踊り字(諾々)、促音便、撥音便、連濁(お堀端)、代用字(表外文字を常用漢字に置き換える。坐→座)、誤用、交ぜ書き(埠頭→ふ頭)、表記ゆれ、当て字(意味に関係なく漢字の音を当てた語句、例:出鱈目)、熟字訓(漢字の音に関係なく熟語の意味で読ませる語句、例:明後日)

言語変種:

声調または高低アクセントの型:

特性: 漢字表現、当て字(ateji representation)、異綴語 spelling variant(Q115819543)

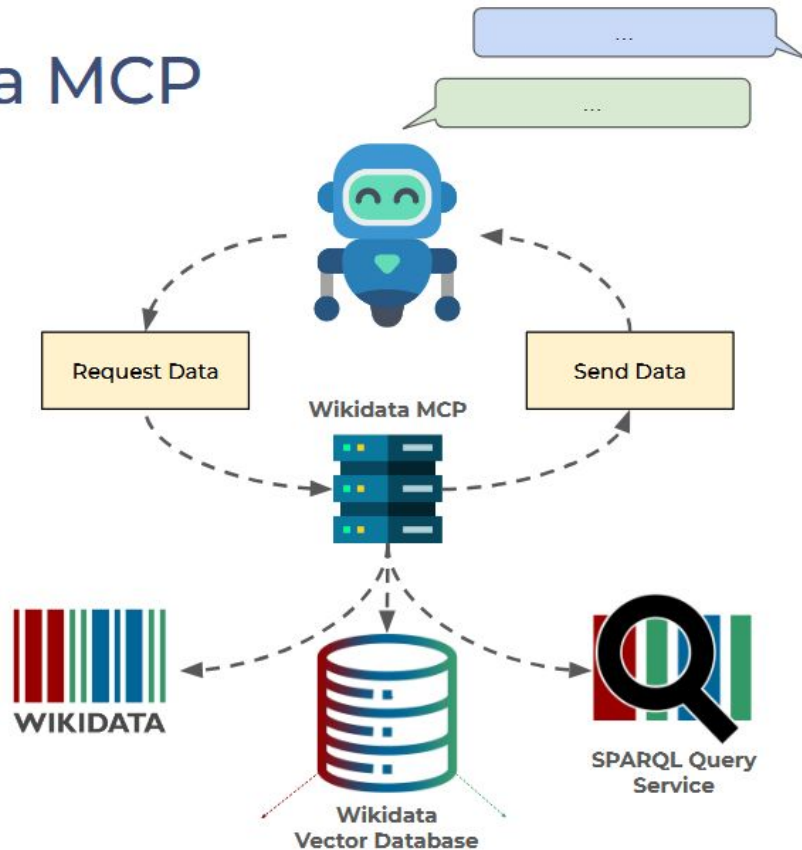
読みの種類: 訓読み、音読み

異綴語(P8530): 語形の表記ゆれへのリンク

当てはまる役割(P3831): 表記ゆれ

数詞(例:手紙に対する「通」)

Wikidata MCP

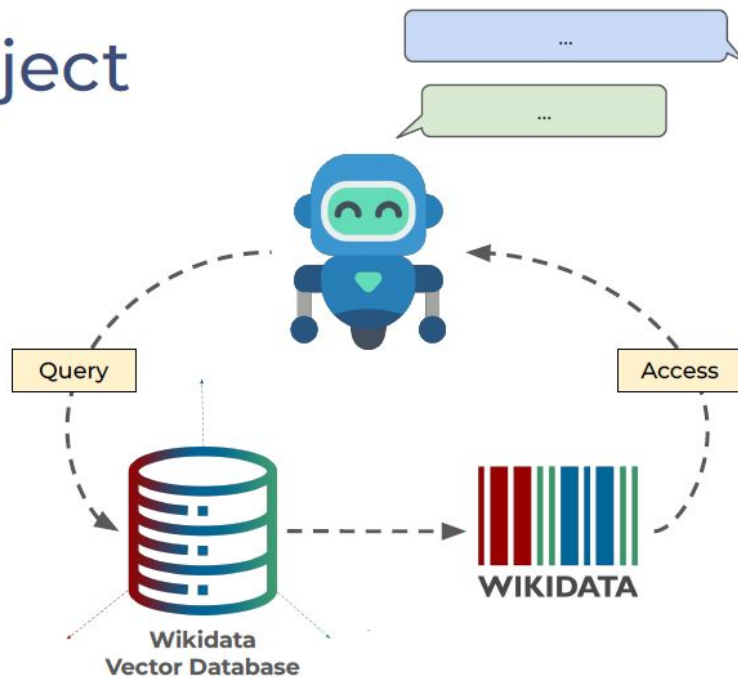


WikidataCon 2025
Oct 31 - Nov 02

Wikidata Embedding Project

Goals

- Creating an open-source Vector Database on top of Wikidata.
- Improve accessibility by enabling semantic search on Wikidata.
- Encourage the open-source AI/ML community to build AI projects with Wikidata.



WikidataCon 2025

Oct 31 - Nov 02

関連情報

< 語彙素 >

- ・語彙データトップページ https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/ja
- ・全般的な解説と登録操作説明 https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation/ja
- ・日本語モデル解説 https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation/Languages/ja/modeling
- ・関連ツール https://www.wikidata.org/wiki/Wikidata:Tools/Lxicographical_data/ja

< ウィキファンクションズ >

- ・ポータル: https://www.wikifunctions.org/wiki/Wikifunctions:Main_Page

< 抽象ウィキペディア >

- ・ポータル: https://meta.wikimedia.org/wiki/Abstract_Wikipedia/ja
- ・WikidataCon 2025 資料: https://commons.wikimedia.org/wiki/File:WikidataCon_2025_-_Wikidata_and_Abstract_Wikipedia_the_Now_and_the_Future.pdf

< Wikidata Embedding Project >

- ・Embedding Project: https://www.wikidata.org/wiki/Wikidata:Embedding_Project
- ・MCP: <https://www.wikidata.org/wiki/Wikidata:MCP>

クレジット

ご注意:本資料にはCC0を適用しておりますが、一部それ以外の利用ライセンスのコンテンツを含む場合があります。その場合は個々の利用ライセンスに従ってご利用ください。



画像のクレジット



[Nikki](#)
CC0



[Open Knowledge foundation](#)
CC BY 4.0