



# Wikidata Lexemes in Brief

Fariz Darari, Ph.D. (Fasilkom UI)



# Context

# /di.le.ma/

## Diskusi Leksem Bersama

Pemberitahuan Hasil Riset, Konsensus Penerjemahan,  
dan Perkembangan Terkini Proyek PPLL



**Ivan Lanin**

Direktur  
Narabahasa

**Benny**

Kontributor Wikidata  
Leksem bahasa Indonesia

**Taufik Rosman**

Kontributor Wikidata  
Leksem bahasa Melayu

**Felicia N. U.**


Konsultan  
Bahasa

**Fariz Darari**

Dosen Fasilkom  
Universitas Indonesia



 **Pendaftaran**  
[bit.ly/wmid-dilema-0323](https://bit.ly/wmid-dilema-0323)

 **Sabtu, 18 Maret 2023**  
14.00 - 16.00 WIB

 Daring via **Zoom**

**Narahubung**  Kartika (0851-9500-2910)  [ppll@wikimedia.or.id](mailto:ppll@wikimedia.or.id)  @wikidataid

# Wikidata Lexemes

## L-id Lexeme

*a meaning carrying part of a language, such as a word or a phrase*

one

### Lemma

*standard form or dictionary form of the lexeme  
for verbs this is usually the infinitive form, for a noun the nominative singular, etc.*

one

### Lexical category

*also known as the part of speech or word class  
defines the lexeme to be either a noun, or a verb, or an adjective, etc.  
the set of possible values is open and taken from the Wikidata items*

one

### Language

*chosen in an open list from Wikidata items*

many

### Statements

*e.g. derived-from, region, period, usage example, homonym, etc.*

## Forms

*specific, conjugated or inflexed forms of the lexeme*

one

### Representation

*the actual string value realizing a given form*

many

many

### Grammatical features

*e.g. normative, comparative, past tense, etc.*

many

many

### Statements

*e.g. region, period, pronunciation, etc.*

many

one

## Senses

### Gloss

*short description, translatable in all languages of Wikidata*

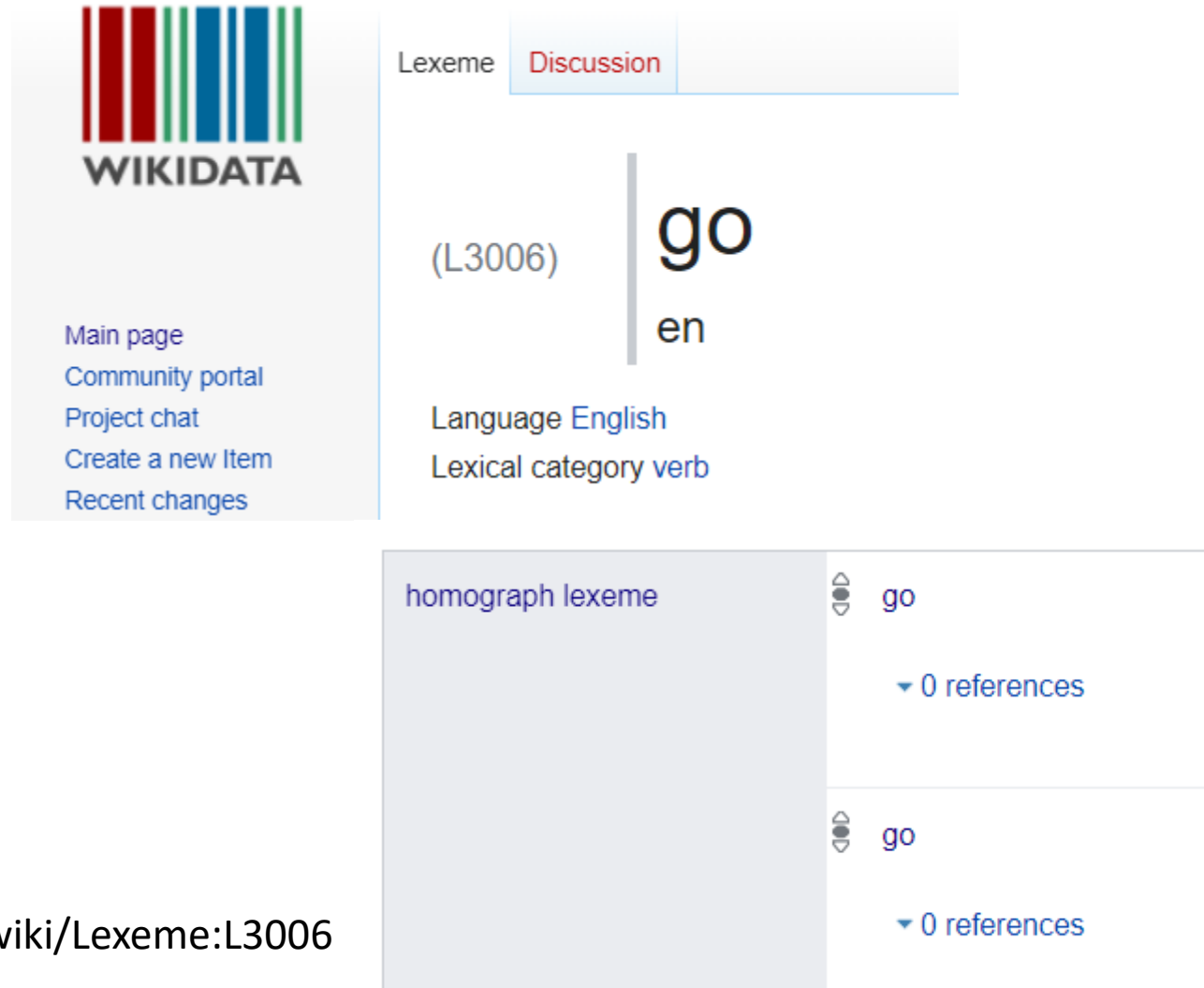
many

many

### Statements

*e.g. translations, synonyms, connotation, register, refers-to-concept, etc.*

# An Example of a Lexeme



The image shows a screenshot of the Wikidata interface for the Lexeme L3006. On the left is the Wikidata logo and a sidebar with navigation links: Main page, Community portal, Project chat, Create a new Item, and Recent changes. The main content area has a 'Lexeme' tab with a sub-tab 'Discussion'. Below this, the identifier '(L3006)' is shown next to a vertical line that separates the word 'go' (top) from its language code 'en' (bottom). Further down, it specifies 'Language English' and 'Lexical category verb'. At the bottom, there are two entries for 'homograph lexeme', each with a dropdown arrow, the word 'go', and a link to '0 references'.

<https://www.wikidata.org/wiki/Lexeme:L3006>

# An Example of a Lexeme

## Forms

L3006-F1 | go  
en

Grammatical features `simple present`

Statements about L3006-F1

L3006-F2 | goes  
en

Grammatical features `third person, singular, simple present`



Statements about L3006-F2


# An Example of a Lexeme

## Senses

	English	move; change location	 edit
	Russian	идти	
	Dutch	gaan	
	Spanish	ir	
	Chinese (Taiwan)	走	
	Traditional Chinese	走	
	Chinese	走	
	Min Nan Chinese	kiáⁿ	
L3006-S1	Gujarati	પસવું; સ્થળ બદલવું	
	German	sich fortbewegen; gehen	
	Italian	andare	
	Kannada	ಹೋಗು; ಹೋಗಿ	
	Hindi	जाना	
	Croatian	ići	
	French	aller	
	Georgian	წასვლა	
	Turkish	gitmek	

## Statements about L3006-S1

**antonym**  stay (English) - remain in a given condition; continue  edit

**translation**  andare (Italian) - go

▼ 0 references

# Queries

- A query is asked to retrieve knowledge about some topic (for instance, about lexemes).
- In knowledge graphs (e.g., Wikidata), queries are asked using a language called: SPARQL
- In the following, we will see a list of SPARQL queries along with their natural language description.



## Query 1: Indonesian nouns starting with "m"

```
SELECT DISTINCT ?l ?lemma WHERE {  
  ?l a ontolex:LexicalEntry ;  
    dct:language wd:Q9240 ;  
    wikibase:lemma ?lemma ;  
    wikibase:lexicalCategory wd:Q1084 .  
  FILTER regex (?lemma, "^m") .  
}
```



# Query 1: Indonesian nouns starting with "m"

Q wd:L693001

monumen

Q wd:L696827

minuman

Q wd:L696816

makanan

• • • • •

## Query 2: Top-5 longest Indonesian nouns\*

```
SELECT DISTINCT ?l ?lemma ?len WHERE {  
  ?l a ontolex:LexicalEntry ;  
    dct:language wd:Q9240 ;  
    wikibase:lemma ?lemma ;  
    wikibase:lexicalCategory wd:Q1084 .  
  BIND(STRLEN(?lemma) AS ?len)  
  FILTER(!(CONTAINS(?lemma, " ")) &&  
    !(CONTAINS(?lemma, "-")))  
} ORDER BY DESC(?len) LIMIT 5
```

\*The noun must not contain whitespace or hyphen.

## Query 2: Top-5 longest Indonesian nouns\*

I	lemma	len
<a href="#">Q wd:L694883</a>	penginternasionalan	19
<a href="#">Q wd:L699305</a>	pertelekomunikasian	19
<a href="#">Q wd:L694882</a>	pengintensifikasian	19
<a href="#">Q wd:L694715</a>	pendiversifikasian	18
<a href="#">Q wd:L699808</a>	pertanggungjawaban	18

\*The noun must not contain whitespace or hyphen.

## Query 3: Senses for noun "anak" (Indonesian)

```
SELECT DISTINCT ?sense ?definition WHERE {  
  ?l a ontolex:LexicalEntry ;  
    dct:language wd:Q9240 ;  
    wikibase:lemma "anak"@id ;  
    wikibase:lexicalCategory wd:Q1084 ;  
    ontolex:sense ?sense .  
  ?sense skos:definition ?definition .  
}
```

## Query 3: Senses for noun "anak" (Indonesian)

sense	↕	definition
<a href="#">wd:L489259-S1</a>		keturunan tingkat pertama
<a href="#">wd:L489259-S2</a>		manusia yang masih kecil

**Thanks!**