Wikidata Lexemes in Brief

Fariz Darari, Ph.D. (Fasilkom UI)

Context

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

many **one** **Representation**
the actual string value realizing a given form

many **Grammatical features**
e.g. normative, comparative, past tense, etc.

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

Senses

many **one** **Gloss**
short description, translatable in all languages of Wikidata

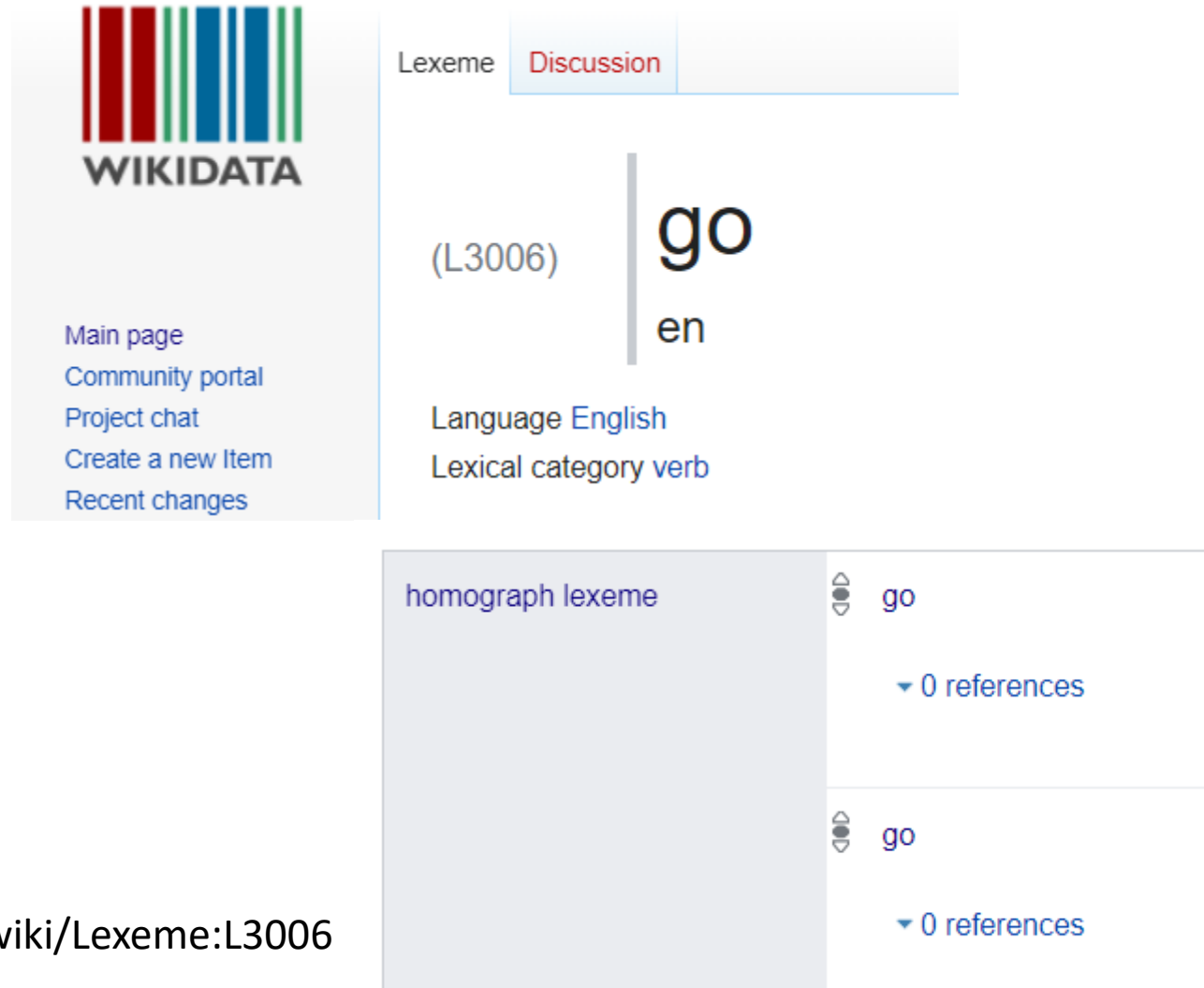**many** **Statements**
e.g. translations, synonyms, connotation, register, refers-to-concept, etc.

An Example of a Lexeme

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	kiâⁿ	
L3006-S1	Gujarati	ખસવું; 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"

🔍 wd:L693001	monumen
🔍 wd:L696827	minuman
🔍 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*

l		lemma	len
🔍 wd:L694883		penginternasionalan	19
🔍 wd:L699305		pertelekomunikasian	19
🔍 wd:L694882		pengintensifikasian	19
🔍 wd:L694715		pendiversifikasian	18
🔍 wd:L699808		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
wd:L489259-S1	keturunan tingkat pertama
wd:L489259-S2	manusia yang masih kecil

Thanks!