

விக்஑ித்தரவு

-நீச்சல்காரன்
நிரலாளர், விக்஑ிப்பீடியர், எழுத்தாளர்

neechalkaran@gmail.com
www.neechalkaran.com
<http://github.com/Neechalkaran>
<http://facebook.com/neechalkaran>

விக்கித் திட்டங்கள்

- ❖ விக்கிப்பீடியா - கட்டுரைகள்
- ❖ விக்கிசனரி - அகரமுதலி
- ❖ விக்கிமூலம் - நூலகம்
- ❖ விக்கிநூல் - புத்தகம்
- ❖ பொதுவகம் - பல்லாடகம்
- ❖ விக்கித்தரவு - தரவு



இடர்

- ❖ தரவு புதுப்பித்தல்
- ❖ மீள் தொடர் செயல்
- ❖ தரவுகளைத் திரட்டல்
- ❖ ஒருங்கிணைந்த தளம்
- ❖ எந்திரவழி கற்றலின் சிக்கல்



ARTIFICIAL INTELLIGENCE

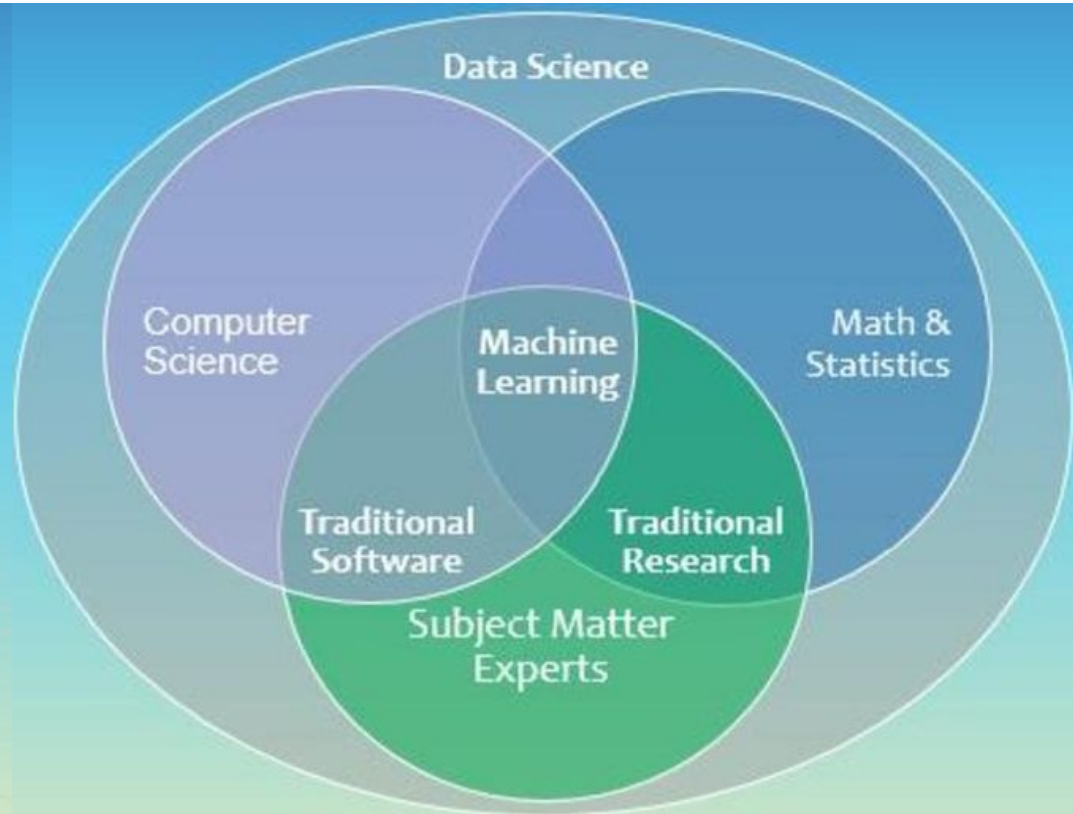
A program that can sense, reason, act, and adapt

MACHINE LEARNING

Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING

Subset of machine learning in which multilayered neural networks learn from vast amounts of data



விக்சித்தரவு

- ❖ 8,76,97,722-> 8.76 கோடி தரவுகள்
- ❖ ஒன்றோடு பிணைப்பு
- ❖ மனிதன் + எந்திரம் கற்றல்
- ❖ Wikibase
- ❖ Wiki statement, Quarry, SPARQL, API
- ❖ உலக அறிவு + மொழி அறிவு

Basava Sagar Dam (Q4866284)

dam in India

 [edit](#)



▼ [In more languages](#)

Language	Label	Description	Also known as
English	Basava Sagar Dam	dam in India	
Tamil	பசுவ சாகர் அணை	No description defined	

[All entered languages](#)

Statements

instance of	 dam	 edit
	▶ 1 reference	
	+ add value	

country	 India	 edit
	▼ 0 references	
	+ add reference	

- [in page](#)
- [community portal](#)
- [subject chat](#)
- [create a new item](#)
- [create a new lexeme](#)
- [recent changes](#)
- [random item](#)
- [query service](#)
- [nearby](#)
- [print](#)
- [help](#)
- [about](#)
- [contact](#)
- [tools](#)
- [what links here](#)
- [recent changes](#)
- [special pages](#)
- [permanent link](#)
- [page information](#)
- [concept URI](#)
- [edit this page](#)

அமைப்பு

- ❖ உருப்படி → பண்பு → மதிப்பு
 - ❖ Earth(Q2)->highestpoint(P610)->Mt Everest(Q513)
 - ❖ Mt Everest(Q513)->elevation above sealevel(P2044) ->8848mt

❖ **Item** - topic

❖ **Property** - how data relevant to this item

❖ **Value** - another Item/value



பெயர்வெளி

-1	Special	Special:SpecialPages	MediaWiki automatic generated content. You cannot edit these.		
0	<i>(None)</i>	Q1	<i>Main name space, for items.</i>	1	Talk:
2	User:	User:userid	User pages	3	User talk:
4	Wikidata: WD: Project:	Wikidata:Glossary	Policies, guidelines, portals, wikiproject pages.	5	Wikidata talk: WT: Project talk:
6	File:		Media description pages (currently not used)	7	File talk:
8	MediaWiki:		User interface customization. Protected by default.	9	MediaWiki talk:
10	Template:	Template:Welcome	Templates	11	Template talk:
12	Help:	Help:Contents	Help pages	13	Help talk:
14	Category:		Category description pages	15	Category talk:
120	<i>Property:</i>	Property:P17	<i>Property definition pages</i>	121	Property talk:
122	<i>Query:</i>	Query:U1	<i>Query definition pages (currently unused)</i>	123	Query talk:
146	<i>Lexeme:</i>	Lexeme:L1	Lexemes	147	Lexeme talk:
640	EntitySchema:	EntitySchema	Schemas	641	EntitySchema talk:
828	Module:	Module:Wikidata	Scribunto modules	829	Module talk:
1198	Translations:	Translations:Wikidata:Introduction/2/en	Used by the translation interface	1199	Translations talk:
2300	Gadget:		Not currently in use	2301	Gadget talk:
2302	Gadget definition:		Not currently in use	2303	Gadget definition talk:
2600	Topic:		Structured Discussions topics	No associated discussion namespace	

Statement

- ❖ `{{#statements:p364}}<!--கொடுக்கப்பட்ட பக்கத்தின் 364 பண்பைக் காட்டும் -->`
- ❖ `{{#statements:p364|from=Q7532310}}<!--7532310 உருப்படியின் 364 பண்பைக் காட்டும் -->`
- ❖ `{{#property:p35|from={{Indian_States_Wikidata_QId|தமிழ்நாடு}}}} #தமிழ்நாட்டின் ஆளுநர்பெயர்கொடுக்கும்`

API

- ❖ தலைப்பிலிருந்து உருப்படி எண்
- ❖ https://en.wikipedia.org/w/api.php?action=query&prop=pageprops&ppprop=wikibase_item&redirects=1&titles=Hindu Religious and Charitable Endowments Department
- ❖ உருப்படி எண்ணிலிருந்து தலைப்பு
- ❖ <https://www.wikidata.org/w/api.php?action=wbgetentities&props=labels&ids=Q12974460&languages=ta&format=xml>
- ❖ <https://www.wikidata.org/w/api.php>
- ❖ <https://colab.research.google.com/drive/1QT9k3PCdUkN9oRyeKHCL5ocvdbyXWtd2>

Wikibase Schema

Term store

wbitem_terms

- 🔑 wbit_id BIGINT
- ◆ wbit_item_id INT
- ◆ wbit_term_in_lang_id INT

wbt_term_in_lang

- 🔑 wbtl_id INT
- ◆ wbtl_type_id INT
- ◆ wbtl_text_in_lang_id INT

wbt_text_in_lang

- 🔑 wbxl_id INT
- ◆ wbxl_language VARBINARY(20)
- ◆ wbxl_text_id INT

wbt_type

- 🔑 wby_id INT
- ◆ wby_name VARBINARY(45)

wbproperty_terms

- 🔑 wbpt_id INT
- ◆ wbpt_property_id INT
- ◆ wbpt_term_in_lang_id INT

wbt_text

- 🔑 wbx_id INT
- ◆ wbx_text VARBINARY(255)

Quarry

- ❖ <https://quarry.wmflabs.org/>
- ❖ எளிய SQL நிரல்
- ❖ Wikidatawiki_p - database
- ❖ page, wbt_item_terms, wbt_property_terms
- ❖ <https://quarry.wmflabs.org/query/46337> Get label from id
- ❖ <https://quarry.wmflabs.org/query/46338> Get wikidata id from label

SPARQL

- ❖ <https://en.wikibooks.org/wiki/SPARQL>
- ❖ <https://query.wikidata.org/>

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, homonym, etc.

Forms

specific, conjugated or inflected forms of the lexeme

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

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

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

விக்கித்தரவில் உள்ளிடும் முக்கியக் கூறுகள்

இலக்கணம்

Case/வேற்றுமை

Direct/Nominative/எழுவாய்

1

Accusative/செயப்படுபொருள்

2

Instrumental/Ergative/கருவி/உடனிகழ்ச்சி

3

Dative/கொள்ளும்/கொடை

4

Ablative/பொருண்மை/நீங்கல்

5

Genitive/Possessive/உடைமை

6

Locative\இட

7

vocative\விளி

8

Terminative	வரை	நாளாவரை
Prolative	வழியாக	கடல்வழியாக
Ornative	கொண்ட	கதவு கொண்ட வீடு
Comparative/Equative	போல	மன்னர் போல மக்கள்

https://en.wikipedia.org/wiki/Template:Grammatical_cases

Lexeme



- Main page
- Community portal
- Project chat
- Create a new Item
- Create a new Lexeme
- Recent changes
- Random Item
- Query Service
- Nearby
- Help
- Donate
- Tools
- Special pages
- Printable version

Special page

Create a new Lexeme

By clicking "Create", you agree to the [terms of use](#), and you irrevocably agree to release y

Create a new Lexeme

Lemma

Language of Lexeme

Lexical category

Create

English

L303548-F1 வீடு ta

Grammatical features singular

Statements about L303548-F1

L303548-F3

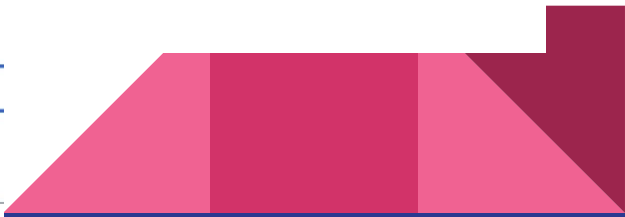
Representation ×

Spelling variant

Grammatical features

dative case ×

Statements about L303548-F3



Text to lexemes

அம்மா ஒரு நாள் வீட்டிற்கு வந்தார்

Language:

தமிழ் - ta - Tamil

Submit

Extraction

Word	Form	Lexeme	Lexical category
அம்மா	அம்மா	அம்மா	noun
ஒரு	ஒரு	எண் பலகை	noun

கருவிகள்



<https://tools.wmflabs.org/ordia/>



<https://wikidata-game.toolforge.org/distributed/#game=49&opt=%7B%22type%22%3A%22horse%22%7D>



<https://lucaswerkmeister.github.io/wikidata-lexeme-graph-builder/>



<https://reasonator.toolforge.org/?q=Q1001>



<http://prowd-prototype.herokuapp.com/profile/110>



<https://github.com/Wikidata/StrepHit>



மேற்கோள்கள்

<https://www.wikidata.org/wiki/Help:FAQ>

<https://books.google.co.in/books?id=2vpRCgAAQBAJ>

<https://blog.wikimedia.de/2015/02/23/platypus-a-speaking-interface-for-wikidata/>

<http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/44818.pdf>



#genre(P136) = Poetry(Q482) && instance(P31) = literary work(Q7725634) && language of work(P407) = Tamil(Q5885)

```
SELECT ?l ?lLabel WHERE {
```

```
  ?l wdt:P136 wd:Q482.
```

```
  ?l wdt:P31 wd:Q7725634.
```

```
  ?l wdt:P407 wd:Q5885.
```

```
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ta". }
```

```
}
```

#Lexemes in Tamil that match an expression

```
select ?lexemeld ?lemma WHERE {
```

```
  ?lexemeld <http://purl.org/dc/terms/language> wd:Q5885;
```

```
  wikibase:lemma ?lemma.
```

```
  FILTER (regex(?lemma, '^ந.*'))# only those lemmas that begin with "ந", i.e. "நாள்"
```

```
}
```

#instance(P31) = literary work(Q7725634) && (language of work(P407) = Tamil(Q5885) || language of work(P407) = Sanskrit(Q5885))

```
SELECT ?l ?lLabel WHERE {
```

```
  ?l wdt:P31 wd:Q7725634.
```

```
  { ?l wdt:P407 wd:Q11059 } UNION { ?l wdt:P407 wd:Q5885 }
```

```
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ta". }
```

நன்றி

Username: Neechalkaran
neechalkaran@gmail.com

<http://github.com/neechalkaran>

<http://twitter.com/neechalkaran>

<http://facebook.com/neechalkaran>

www.neechalkaran.com