

Wikifunctions and Abstract Wikipedia: Community Perspectives



Houcemeddine Turki

Data Engineering and Semantics Research Unit, Faculty of Sciences of Sfax, University of Sfax, Tunisia

Wikimedia Tunisia, Tunisia

Wikimedia and Libraries User Group, United States of America





About Us

General Introduction

Located in Tunisia (270 km from Tunis)

One of the leading public universities in Tunisia

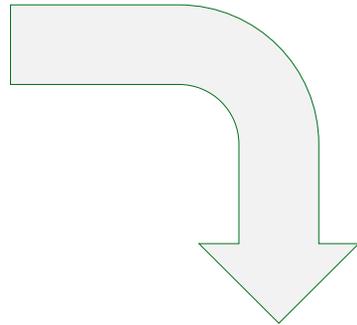
One of the leading African universities in Computer Science Research





- » Recently Created Research Unit at the Faculty of Sciences of Sfax, University of Sfax
- » Specialized in many areas of Computer Science ranging from Semantic Technologies, Social Network Analysis and Graph Embeddings to Library and Information Science
- » One major achievement is the development of computer applications driven by Wikimedia Projects for sustaining digitization in the Arab countries and in Africa
- » Published in highly referred scholarly journals
- » Established a formal collaboration with the Wikimedia Community through joint projects with Wikimedia Tunisia

Abstract Wikipedia



- » The Project began on May 22, 2020 after its approval by the Board of Trustees of Wikimedia Foundation.
- » Aims to reduce the concerns about the lack of coverage of many topics in most of the Wikimedia Projects, particularly Wikipedia.
- » Part of the ninth point of Wikimedia 2030 Strategy: Innovation in Free Knowledge
- » Tries to create a platform for the automatic creation of articles about Wikidata items in 2023.



Team



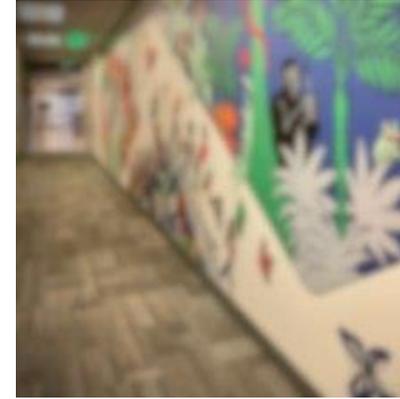
Denny Vrandečić
Project Director



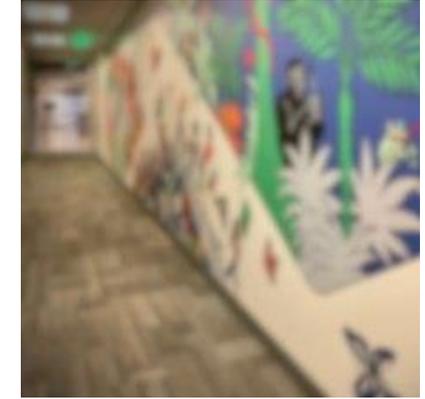
Lukas Werkmeister
Senior Developer



Adam Beso
Engineering Director



Cory Massaro
Staff Software Engineer



Genoveva Galarza
Staff Software Engineer



James Forrester
Staff Software Engineer



Aishwarya Vardhana
Senior UX Designer



Cai Blanton
Senior Engineering Manager



Nick Wilson
Community Relations Specialist



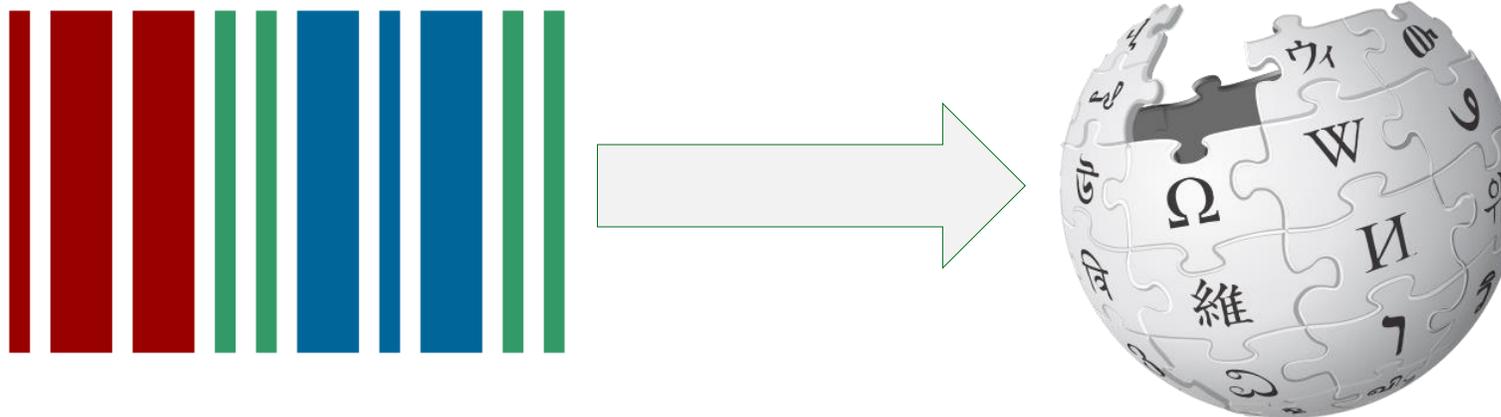
Introduction

How Wikifunctions and Abstract
Wikipedia have evolved

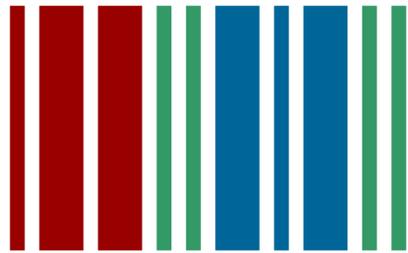
However...

The Arabic Community cannot create
19 million articles by hand from
scratch

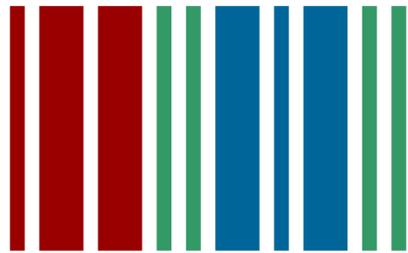
Using the semantic information about the **Wikidata items** covering all Wikipedia articles to create articles about entities not supported by the **Arabic Wikipedia**



Abstract Wikipedia



Names of Wikidata Entities
in Natural Languages and
Wikidata Statements



Lexical Database for Natural
Languages: Lexemes, Forms
and Senses



Database for Linguistic
Rules



Abstract Wikipedia

Names of Wikidata Entities in Natural Languages and Wikidata Statements

WikiArabia (Q4994929)

Arabic-speaking Wikimedians conference

 edit

▼ In more languages

Configure

Language	Label	Description	Also known as
English	WikiArabia	Arabic-speaking Wikimedians conference	
French	WikiArabia	conférence wikimédienne arabophone	
Russian	No label defined	No description defined	
Tunisian Arabic (Arabic script)	No label defined	No description defined	
Arabic	ويكي عربية	مؤتمر الويكيبيديين الناطقين باللغة العربية	
German	WikiArabia	arabische Wikimedia-Konferenz	

[Fewer languages](#)

Statements

instance of	 convention	 edit
	▼ 0 references	+ add reference
		+ add value

Lexical Database for Natural Languages: Lexemes, Forms and Senses

(L314711)

عَلَّقَ

ar

عَلَّقَ

ar-x-Q775724

 edit

Language [Arabic](#)

Lexical category [verb](#)

Statements

[+ add statement](#)

Senses

L314711-S1

French

faire un commentaire

 edit

English

do a comment

Statements about L314711-S1

[synonym](#)



commenter (do a comment)

 edit

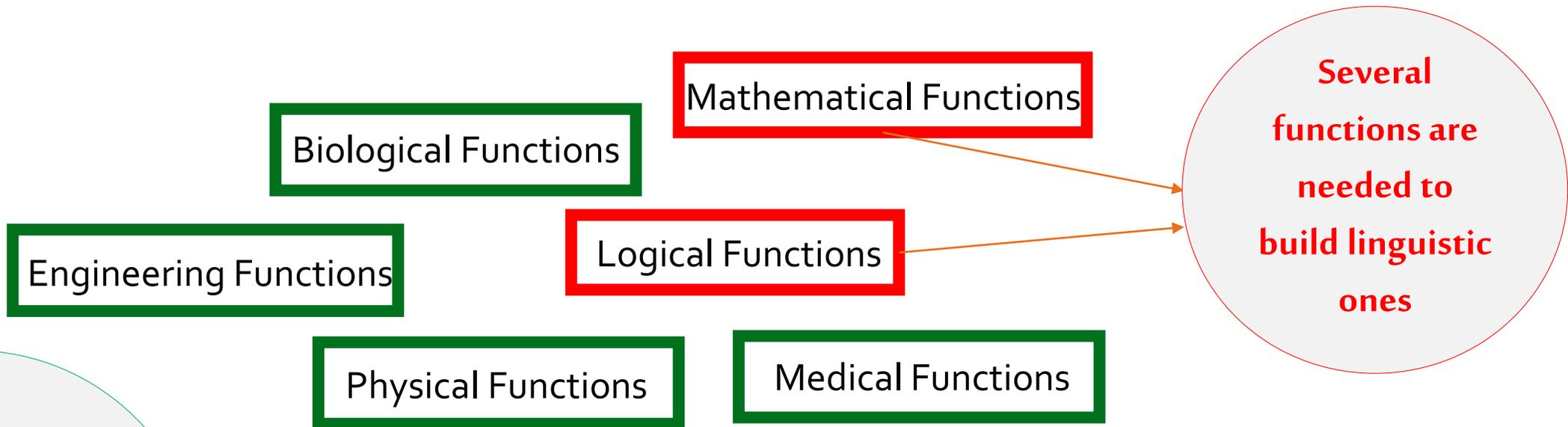
▼ 0 references

[+ add reference](#)





Wikifunctions: More than a simple database for linguistic functions



Several others are useful for driving interesting applications of Wikidata

- Clinical Decision Support
- Validation of Wikidata Statements
- Generation of New Knowledge for Wikidata
- Use in SPARQL Queries
- Use in Wikimedia Tools, Templates, Bots and Scripts

Importance of Linguistic Functions

- Using natural language generation to bootstrap missing Wikipedia articles: A human-centric perspective

Cite

- » Validation and Enrichment of Lexicographical Data and Wikidata Labels in Wikidata
- » Natural Language Generation
- » Machine Translation of Wikipedia Articles
- » By contrast to Machine Learning, the use of functions allows the transparent tracking and troubleshooting of flaws.

Authors: Kaffee, Lucie-Aimée | Vougiouklis, Pavlos | Simperl, Elena

Article Type: Research Article

Abstract: Nowadays natural language generation (NLG) is used in everything from news reporting and chatbots to social media management. Recent advances in machine learning have made it possible to train NLG systems that seek to achieve human-level performance in text writing and summarisation. In this paper, we propose such a system in the context of Wikipedia and evaluate it with Wikipedia readers and editors. Our solution builds upon the ArticlePlaceholder, a tool used in 14 under-resourced Wikipedia language versions, which displays structured data from the Wikidata knowledge base on empty Wikipedia pages. We train a neural network to generate an introductory ... [Show more](#)

Keywords: Wikipedia, Wikidata, ArticlePlaceholder, multilingual, natural language generation, neural networks

DOI: 10.3233/SW-210431

Citation: *Semantic Web*, vol. Pre-press, no. Pre-press, pp. 1-30, 2021

[Get PDF](#) 





- » Beta edition currently available at <https://notwikilambda.toolforge.org/>
- » To create a function, we need three items:
 - » **Function item:** Provides semantic information about the function such as the names and description of the function in multiple natural languages and the names and types of the inputs and the output.
 - » **Implementation item:** Allows to embed source codes in programming languages (Python, JavaScript or Lua) or to develop the function through the association of other open functions.
 - » **Tester item:** Adds a test case to the Function to validate the proposed implementations.

Function item

- » Allows to define the function through:
 - » Adding its names and descriptions in multiple languages
 - » Specifying the names and types of the inputs
 - » Precising the type of the output
- » Does not currently allow to link the Function and its inputs to their equivalents in Wikidata. Such an improvement can be useful to automate the use of the functions.
- » Wikidata items and lexemes cannot be used as inputs. An input should rather be a string, a char, a value and a boolean.

Root and Pattern-Based Generator of Lexemes for Arabic Languages (Z10157)

Function (Z8)

Hide labels box

Language	Label	Also known as
English	Root and Pattern-Based Generator of Lexemes for Arabic Languages	
Danish	rod og mønster-baseret leksemgenerator for arabiske sprog	

Contents

Function

return type

String

Argument List

type: String

label:

English: Root (Letters separated by spaces)

Arabic: رُود

Danish: rod

type: String

label:

English: Pattern

Arabic: مَوْزُون

Danish: mønster



Implementation item

- » Includes an implementation of the function in a programming language: Python, JavaScript or Lua
- » Can instead define an association of other open functions that can be used to return the required output
- » Inputs are called using its identifier as defined in the Function item: First defined input is K1 and so on
- » It is linked to the Function item to enable the running of the implemented function

Implementation for Root and Pattern-Based Generator of Lexemes for Arabic Languages (Z10158)

Implementation (Z14)

Hide labels box

Language	Label	Also known as
English	Implementation for Root and Pattern-Based Generator of Lexemes for Arabic Languages	

Contents

Implementation

function: [Root and Pattern-Based Generator of Lexemes for Arabic Languages](#) (This object is associated with its function.)

```
python
1 def Z10157(Z10157K1, Z10157K2):
2     r = Z10157K1.split(" ")
3     pattern = Z10157K2
4     placeholder = ["ل", "ع", "ف"]
5     word = ""
6     i = 0
7     for k in range(len(pattern)):
8         if (i > 2):
9             if (pattern[k] == placeholder[2]):
10                word += r[i]
11                i += 1
12            else:
13                word += pattern[k]
14        if (i <= 2):
15            if (pattern[k] == placeholder[i]):
16                word += r[i]
17                i += 1
18            else:
19                word += pattern[k]
20    return word
```



Tester item

- » Involves a test value for every input of the function as well as the expected value of the output.
- » It should be linked to the Function item to enable its use for the validation of implementation items.

Tester for Root and Pattern-Based Generator of Lexemes for Arabic Languages (Z10159)

Tester (Z20)

Hide labels box

Language	Label	Also known as
English	Tester for Root and Pattern-Based Generator of Lexemes for Arabic Languages	

Contents

Tester

function: *Root and Pattern-Based Generator of Lexemes for Arabic Languages* (This object is associated with its function.)

call:

function: *Root and Pattern-Based Generator of Lexemes for Arabic Languages*

- Root (Letters separated by spaces): `String` ك ت ب
- Pattern: `String` مَكْتُوبَةٌ

result validation:

function: *String equality*

- second string: `String` مَكْتُوبَةٌ

Bugs exist, Upgrade required: New user interface is being developed for a better experience

Create a new ZObject Help

This form lets you create a new ZObject, which can represent a Function, its Implementation, a Tester for it, or one of various other kinds of ZObjects. You should read the help documentation and community guidance first.

Make sure that you follow community policies, and copyright and patent rules.

Define Function Provide Tests Write Function

Start Function Name Input(s) Output

Name

What is the name of your function? What does it do?

English

Function Name

[See name examples](#)

Description

Aliases

[See alias examples](#)

Progress
25%

Name
Description

```
graph TD; Input[Input] --> Function[Function]; Function --> Output[Output];
```

Proposed Challenges



- » Data models for Lexicographical Data should be adapted to the morphology of every covered language (Accusative-Dative for German Language, Root and Pattern Generation of Lexemes for the Arabic Languages) **(Priority)**
- » System messages for the embedded compiler should be translatable to enable the understanding of flaws by non-English Wikimedia users (Limited Work has already been done by the Computer Science Community in this context) **(Priority)**
- » Region-Based Logical constraints for validating Wikidata statements should be implemented to support the semantic particularities in countries (Please refer to Whose Knowledge's Decolonizing the Internet's Structured Data) **(Priority)**
- » Checking the adaptability of Wikifunctions and Abstract Wikipedia to cover non-Latin scripts (Effect of NFC normalization for the Arabic Script on the Arabic diacritics)
- » Adding references to Function and Implementation items is mandatory for allowing the reuse of Wikifunctions in Trustworthy Artificial Intelligence
- » Guidelines for Language Transcription should be developed for several languages where linguistic norms do not exist

References

- » Vrandečić, D., Ortiz, M., & Schneider, T. (2018). Toward an Abstract Wikipedia. In *Description Logics*.
- » Vrandečić, D. (2020). Architecture for a multilingual Wikipedia. *arXiv preprint arXiv:2004.04733*.
- » Vrandečić, D. (2021). Building a multilingual Wikipedia. *Communications of the ACM*, 64(4), 38–41.
- » Kaffee, L. A., Vougiouklis, P., & Simperl, E. (2021). Using natural language generation to bootstrap missing Wikipedia articles: A human-centric perspective. *Semantic Web*, (Preprint), 1–30.

Image Credits

- » <https://commons.wikimedia.org/wiki/File:Wikifunctions13logoproposal11.png>
- » [https://commons.wikimedia.org/wiki/File:Wikipedia logo puzzle globe spins horizontally and vertically, revealing the contents of all of its puzzle pieces, without background.gif](https://commons.wikimedia.org/wiki/File:Wikipedia_logo_puzzle_globe_spins_horizontally_and_vertically_revealing_the_contents_of_all_of_its_puzzle_pieces_without_background.gif)
- » [https://commons.wikimedia.org/wiki/File:Denny Vrandecic portrait.jpg](https://commons.wikimedia.org/wiki/File:Denny_Vrandecic_portrait.jpg)
- » [https://commons.wikimedia.org/wiki/File:LucasWerkmeister 2019-10a.jpg](https://commons.wikimedia.org/wiki/File:LucasWerkmeister_2019-10a.jpg)
- » [https://commons.wikimedia.org/wiki/File:Adam Baso portrait.jpg](https://commons.wikimedia.org/wiki/File:Adam_Baso_portrait.jpg)
- » [https://commons.wikimedia.org/wiki/File:Wikimedia Foundation hall blurred.jpg](https://commons.wikimedia.org/wiki/File:Wikimedia_Foundation_hall_blurred.jpg)
- » [https://commons.wikimedia.org/wiki/File:Forrester, James March 2018.jpg](https://commons.wikimedia.org/wiki/File:Forrester,_James_March_2018.jpg)
- » [https://commons.wikimedia.org/wiki/File:Wilson, Nick January 2016.jpg](https://commons.wikimedia.org/wiki/File:Wilson,_Nick_January_2016.jpg)
- » [https://commons.wikimedia.org/wiki/File:Wikilambda - Early Eta - Create a new function.webm](https://commons.wikimedia.org/wiki/File:Wikilambda_-_Early_Eta_-_Create_a_new_function.webm)
- » <https://commons.wikimedia.org/wiki/File:Symbol-User-Email-Icon.png>
- » [https://commons.wikimedia.org/wiki/File:Phone communication icon.png](https://commons.wikimedia.org/wiki/File:Phone_communication_icon.png)
- » [https://commons.wikimedia.org/wiki/File:Twitter icon.png](https://commons.wikimedia.org/wiki/File:Twitter_icon.png)



Questions

Questions from WikidataCon 2021

« We can probably have duplicate Wikidata items. How to solve this? »

– Adel Nehaoua

» Current Situation:

- Merging items in Wikidata is done by hand
- Merging items in Wikidata requires a lot of human efforts
- Many duplicate Wikidata items still exist
 - Caused by Article duplication in a particular language edition of Wikipedia
 - Caused by the lack of human recognition of a Wikipedia article in a given language (Example: « Road Surface » [Q1049667], « Chaussée » [Q1418214])

» How to solve this:

- Creating a Wikidata Game for merging Wikidata items
- Proposing a neural network for merging Wikidata items
- Developing functions in Wikifunctions for merging Wikidata items
- Developing logical constraints preventing the creation of duplicate items based on the similarity of their labels and statements
- Wikimedia Deutschland has worked on the topic for two years. Efforts are still needed to enhance the efficiency of methods.
- Wikipedia Categories and Wikilinks can be useful (Please refer to Wikidata Workshop 2021).



Thank You

Questions can be provided in Arabic Languages,
French, English, Italian, German and Maltese



turkiabdelwaheb@hotmail.fr



+21629499418



@Csisc1994

<https://dblp.org/pid/176/1531.html>

https://www.researchgate.net/profile/Houcemeddine_Turki

