

A sunset over a savanna landscape. The sun is low on the horizon, casting a warm, golden glow across the sky and the land. In the foreground, a herd of animals, possibly buffalo or similar large mammals, is grazing in a field of tall grass. The background shows rolling hills and scattered trees under the bright sky.

African Wikimedia Projects Tracker

Refreshing how we see African language projects

A presentation by: **Dumisani Ndubane**
Derick Alangi

[Danijel Mihajlovic](#), [CC BY-SA 4.0](#), via Wikimedia Commons

Why so much interest?

- African languages are dying. ([Victor Oladokun, 2019](#))
- Connectivity in Africa is improving. ([Forbes, 2016](#))
- 45% of the next 1 Billion internet users will come from Africa. ([Visual Capitalist, 2020](#))
- Wikipedia has grown as a global brand, but growth in Africa is not where it should be. ([De Gruyter, 2014](#))



Africa as a region contributes:

2.2%

Editors Across all Wikimedia projects

2.0%

Unique page views across all projects

Starting at the beginning of the trail

- User:Greenman has consistently documented the state of Wikipedia projects in Africa since his [first blog](#) in December 2008.
- User:Greenman presented a talk on [African Language Wikipedia Statistics](#) at the first Wiki Indaba in 2014.
- User:Kipala presented [African language Wikipedias](#) at the second Wiki Indaba in 2017.
- User:Thuvack & User:Jamie_Tubers presented [State of African Wikis](#) at the 4th Wiki Indaba in 2019.



The overall picture from above

- 114 African Language Project Wikis
 - 55 Language Wikipedias
 - 37 Language Wiktionaries
 - 12 Language Wikibooks
 - 10 Language Wikisource

43 of these projects are in the Incubator

- 6432 Active Editors across the 114 Wikis
(Not disambiguated by username)



Rankings of language wikis by Page Count

Top 10 Wikipedias

- Egyptian Arabic (1,359,802)
- Standard Arabic (1,142,116)
- Afrikaans (100,882)
- Malagasy (94,299)
- Swahili (68,159)
- Yoruba (33,491)
- Amharic (15,010)
- Hausa (12,901)
- Zulu (10,441)
- Sepedi (8,444)

Top 10 Wikitionaries

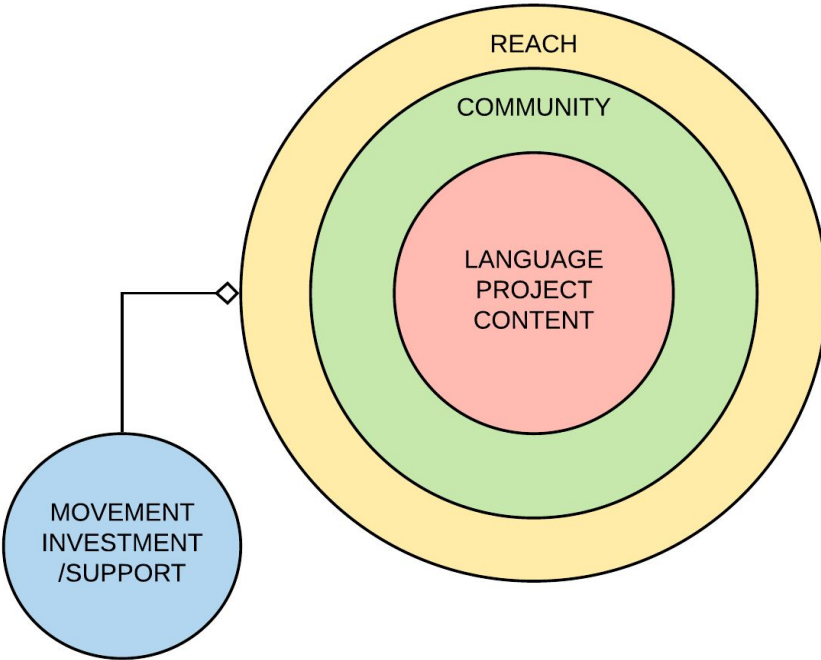
- Malagasy (1,709,212)
- Standard Arabic (66,743)
- Afrikaans (22,026)
- Oromo (18,095)
- Swahili (14,052)
- Wolof (2,574)
- Soomaaliga (1,640)
- Igbo (1,520)
- SeTswana (1,342)
- Bambara (1,138)

**Is that enough to truly
understand the language Wiki
Landscape in the Region?**

Other factors that would be useful

Consider building a language Index using the following factors

| | | | |
|-----|--|------------------------------------|-----|
| 30% | Page Count | Word Count | 20% |
| 10% | Active Editors | Level of Language Localization | 10% |
| 10% | Size of 1 st & 2 nd Speaker population | Language reach by No. of Countries | 10% |
| 10% | Affiliate Support | | |



Building towards a Dashboard

First we need reliable data

Current Wiki data

The media Wiki statistics page is a very good source to mine statistics on Wiki Project content, word count, and active editors.

Python data scraping

Historical Wiki Data

Unfortunately Mediawiki statistics does not provide historical data.

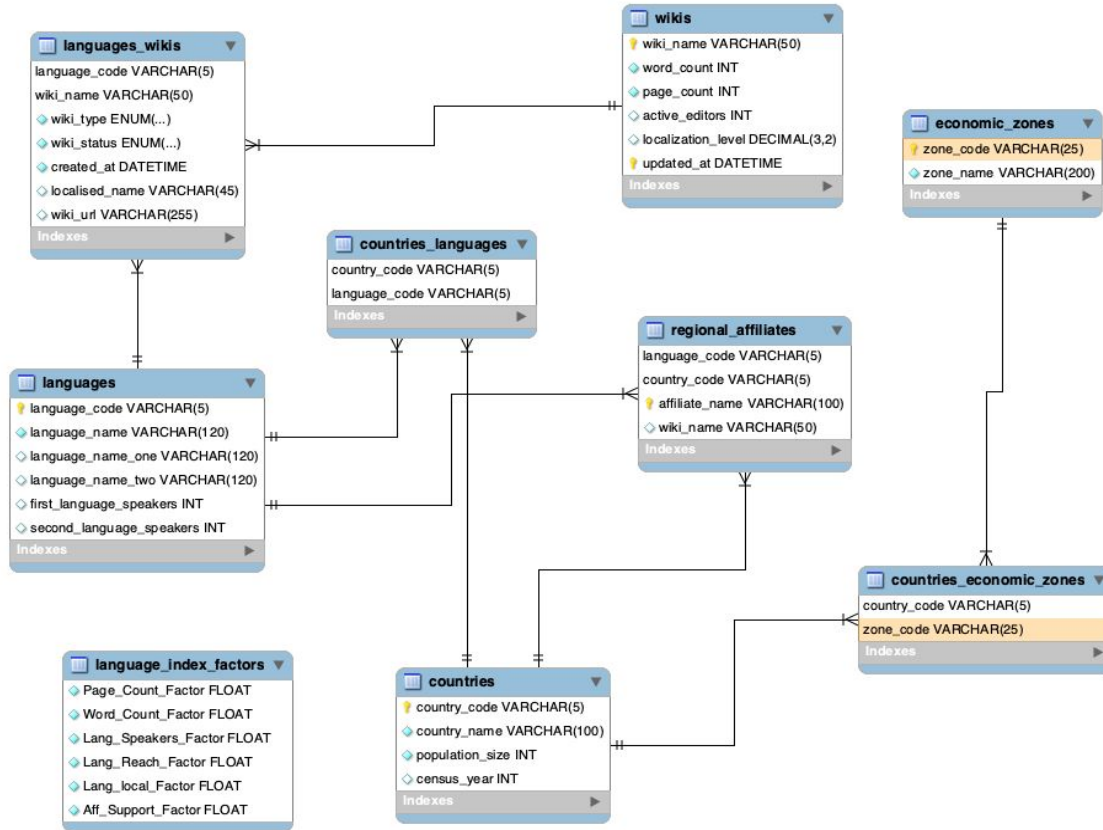
Luckily User:Greenman gives us a place to start with his blogs for Top 10 language wikis by page count.

Exploring using the Internet Archive's Wayback machine to fill-in the blanks.

Language Stats

Many sources available, including wikipedia, providing historical census data breakdowns by language and countries.

Then we need a relational database



The mother tables are
“wikis”, “languages”,
and “countries”

How does the entities
relate with one
another?

Now we are building a Dashboard Tool (Tech Stack)

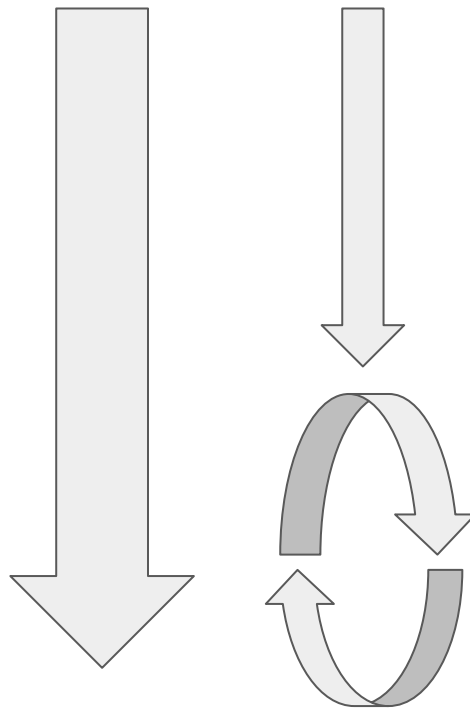


Simple yet flexible JavaScript charting for designers & developers



Now we are building a Dashboard Tool (Plan)

- Development plan - state of the work
 - Design the ER diagram ✓
 - Implement the database on MySQL & import dataset ✓
 - Develop SQL queries for the dashboard filters ✓
 - Setup application environment locally & on toolforge ✓
 - Build and release an MVP (v1.0) dashboard ☰
 - Iterate and add more features ☰
 - Release a stable version of dashboard ☰
 - Collect feedback and improve the tool ☰



Now we are building a Dashboard Tool (Plan)

- Project is open source and anyone can contribute which means:
 - This means you can add code
 - Request for features to be added
 - Contribute to a wiki which will in turn improve the tool ;)

Here is a Mock-up

Rankings of language wikis by Language Index

Top 10 Wikipedias

- Standard Arabic (9.5)
- Egyptian Arabic (6.3)
- Hausa (2.8)
- Fulani (2.8)
- Swahili (2.6)
- Yoruba (2.2)
- Igbo (1.8)
- Afrikaans (1.8)
- Malagasy (1.6)
- Moroccan Arabic (1.5)

We have a long way to go