Wikipedias in African languages 2017

An evaluation based on Wikimedia statistics

Ingo Koll a.k.a. mtumiaji:kipala on sw.wikipedia.org

Kipala: Wikipedias in African languages, 20.01.2017
Greetings from swwiki
Salamu kutoka Wikipedia ya Kiswahili

• Swwiki has been developed by many, in the core there is a team of 5 volunteers who have remained since 10 years
• After wikiindaba 2014 we had a first workshop in Tanzania, thanks to WMF support
• First time our longtime editors could meet
• Important for our motivation to continue
• Phantastic response from teachers and students at Morogoro
• Several 100 edits out of it, some continue
Impression from the Morogoro workshop
Africa – the continent with many languages

2000 – 3000 languages (of 5-7000 worldwide)

But they are threatened
50-90 % estimated to die out until 2100

Any point to start a wikipedia with a “small language” ?

12-15 languages understood by 80% of all Africans in interethnic communication

Concentrate efforts here ?
African Languages are important

- The education systems in foreign languages limit severely the reach of education and spread of knowledge in society.
- Children start their way into learning best in mother tongue (first language).
- Education systems have historically worked with external languages with fair results but limited reach (Latin in Europe, Arabic in Middle East, etc) – but education for all and development came after switch to local languages.
- Countries in Africa have to develop their own languages, but some will not manage because of too much linguistic diversity.
- Wikipedia and open media is among the major chances towards building a knowledge resource in African languages – (under some conditions).
Wikipedia: 295 languages
c.a. 40 African language wikipedias

A look at all wikipedias by size

Huge Wikipedias (more than 1 Mio articles) > 13

Large Wikipedias (100,000 to 1 Mio articles) > 45

Medium size Wikipedias (10,000 to less than 100k) > 77

Small Wikipedias (1,000 to less than 10 k articles) >112

Tiny Wikipedias (100 to less than 1,000 articles) > 38

“Micropedias” (less than 100 articles) > 10
### Wikipedias in African languages 2017:

„African languages“ = languages native to Africa = including Afrikaans + Masri = excluding English, French, Arabic

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huge Wikipedias (1 mio+)</td>
<td>None</td>
</tr>
<tr>
<td>Large Wikipedias (100k+)</td>
<td>None</td>
</tr>
<tr>
<td>Medium size Wikipedias (10k+)</td>
<td>6</td>
</tr>
<tr>
<td>Small Wikipedias (1k+)</td>
<td>10 (+3)</td>
</tr>
<tr>
<td>Tiny Wikipedias (100+)</td>
<td>16 (-7)</td>
</tr>
<tr>
<td>Closed Micropedias</td>
<td>4</td>
</tr>
</tbody>
</table>

(change over 2 yrs)
The African Language Wikipedias

Medium & small size

Minipedias under 1,000 articles
- Zulu
- Oromo
- Tswana
- Xhosa
- Twi
- Tumbuka
- Kirundi
- Swati
- Bambara
- Tsonga
- Sesotho
- Chichewa
- Ewe
- Akan
- Venda
- Tigrinya

Masri 16,200 +3k
Amharic 13,300
Swahili 35,000 +8k
Malagasy 83,200 +36k
Nigerian 31,500
Yoruba 2,800
Kabyle 2,800
Lingala 2,800
Shona 2,700
Kinyarwanda 1,800
Somali 4,300 +1k
Kikuyu 1,300
Kiganda 1,100
(+ change over 2 yrs)
Large community
Large numbers

Example
Egyptian Arabic Masri
76 Mio speakers ±2.5 Mio in Dec 2016

Smaller communities
Lower numbers
Afrikaans
12 Mio speakers
1.4 Mio views in Dec 2017

Pageviews: readers

Number of pageviews
Share of pageviews

not necessarily viewed by many in the community
Masri views in Egypt 1.7% of all wikipedia views there

Afrikaans views reach more than 10% in the language community
Abolute pageviews in ALWs (December 2016)

All ALWs have very small view numbers compared to the huge Wikipedias like English (9,868 Mio) or French (741 Mio)

- Egyptian Masri 2.5 Mio
- Afrikaans 1.4 Mio
- Swahili 1.1 Mio
- Somali 448 k
- Amharic 446 k
- Yoruba 299 k
- Malagasy 283 k
<table>
<thead>
<tr>
<th>ALWs shares per country Dec 2016</th>
<th>Number</th>
<th>% of all Wikipedia views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egyptian Arabic Masri</td>
<td>709.000</td>
<td>1.7%</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afrikaans Wp</td>
<td>334.000</td>
<td>0.9%</td>
</tr>
<tr>
<td>Tanzania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swahili Wp</td>
<td>417.000</td>
<td>6.8%</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swahili Wp</td>
<td>136.000</td>
<td>0.9%</td>
</tr>
<tr>
<td>Nigeria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Igbo Wp</td>
<td>363.000</td>
<td>1.2%</td>
</tr>
<tr>
<td>Yoruba</td>
<td>??</td>
<td>??</td>
</tr>
<tr>
<td>Ethiopia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amharic Wp</td>
<td>108.000</td>
<td>2.6%</td>
</tr>
<tr>
<td>Somalia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somali Wp</td>
<td>91.000</td>
<td>9.1%</td>
</tr>
</tbody>
</table>

Assessing relevance / usefulness of an ALW market shares in “home countries” (Percent of all Wikipedia views in the country)
Competetiveness: Swahili compared to English on local topics

Looking at topics where we can assume few international views in English and that requests on enwiki mostly come from Tanzania. Likely assume that swwiki views come from Tanzania. Swwiki is surprisingly competitve in local views!

Pageviews for selected articles on localities in Tanzania from Jan 2016-Jan 2017

Mbozi Meteorite  en 3.014  sw 3.369
Kigoma Region · en 11.334  sw 8.924
Mbozi District · en 2.177  sw 1.336
Kakonko District en 705  sw 586

(https://tools.wmflabs.org/pageviews) Usually the relation between pageviews is 1 swwiki : 11,000 enwiki
Results of “market share” view

- Most wikipedia views in Africa are in English, French, etc.
- Most ALWs have no traceable views in home country (under 0.1%)
- Somali 9% and Swahili 7-9% have strongest national shares of all ALWs (Somali with small numbers overall)
- Afrikaans has small national but 10% community share!
- Large Malagasy has only 8000 = 0.3% viewers at home
- **Surprise Nigeria:** Yoruba (31,500 art) is not traceable in Nigeria, but Igbo (1,300 art) has 1% of Nigerian views
- Tanzanians open swwiki only a bit less than enwiki on local topics (if article exists already in swwiki; other topics can not be verified)
Reasons for stronger / weaker viewership?

• Probably quality!
• Malagasy & Yoruba have a large share of bot created articles (96% in Malagasy)
• Yoruba has extremely short articles (sometimes not even a sentence), Malagasy just many lists
• Yoruba may have a special problem with its orthography reducing the views via search machines
• Afrikaans and Swahili have followed recommended topics like “1,000” and “10,000 articles every Wikipedia should have”, much local content, bringing views via search machines
• both have hardly any bot-created articles
## Community Interest: Views from home vs. abroad

**Pageviews in ALWs, coming thru IPs from these countries**

<table>
<thead>
<tr>
<th>Language</th>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Igbo (ig)</td>
<td>Nigeria</td>
<td>85.8%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>6.1%</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>U.K.</td>
<td>3.5%</td>
</tr>
<tr>
<td>Swahili (sw)</td>
<td>Tanzania</td>
<td>40.4%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>16.4%</td>
</tr>
<tr>
<td></td>
<td>Kenya</td>
<td>13.2%</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>11.6%</td>
</tr>
<tr>
<td></td>
<td>Ireland</td>
<td>7.4%</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>4.6%</td>
</tr>
<tr>
<td>Afrikaans (af)</td>
<td>South Africa</td>
<td>27.0%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>22.8%</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>18.5%</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>9.4%</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>4.9%</td>
</tr>
<tr>
<td></td>
<td>Netherlands</td>
<td>3.9%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>13.5%</td>
</tr>
<tr>
<td>Somali (so)</td>
<td>Somalia</td>
<td>29.3%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>33.8%</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>12.9%</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>17.4%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>6.6%</td>
</tr>
<tr>
<td>Yoruba (yo)</td>
<td>South Africa</td>
<td>27.0%</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>40.9%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>34.5%</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>13.5%</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>5.6%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>7.6%</td>
</tr>
<tr>
<td>Malagasy (mg)</td>
<td>France</td>
<td>43.1%</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>29.8%</td>
</tr>
<tr>
<td></td>
<td>Germany</td>
<td>14.2%</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>3.7%</td>
</tr>
<tr>
<td></td>
<td>Madagascar</td>
<td>3.7%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>5.5%</td>
</tr>
</tbody>
</table>
Weakness & strength: teams

Wikipedia grew as teamwork
Many ALWs have no or very small teams

In December / October 2016 there were active Wikipedians (who edited 5 times or more):

- In Malagasy 1 / 0
- In Afrikaans 25 / 12
- In Swahili 10 / 3
- In Yoruba 5 / 0
- In Masri 8 / 3
- In Amharic 1 / 1
Observations and assumptions

Perspectives for ALWs
• potential is ahead, internet access grows in Africa.
• Igbo ALW views show strong interest in some communities (measurable views although there is little content)

• The ALWs without team have little perspective

• Quality strategy is more important than quantity – widespread topics bring viewers via searchmachines only if the ALW has topics to offer
• Local content in local language attract views
Challenges for ALWs

- The near monopoly of European languages as medium of higher education
- Weaknesses of most educated Africans to discuss scientific etc. topics in mother tongue & “national language”
- Little faith among Africans in ability of African Languages to express a number of issues
- Reduced ability to read content in ALWs on topics people know through education in foreign languages
Closing remark

- An ALW has good chances for development if its language is medium of instruction at least at primary school level.

- It is not necessary for its development to use a „large“ language (cf the example of Afrikaans).

- A wikipedia in a language which is not medium of instruction at school will likely get stuck on a lower level.