COMMUNITY TRANSLATION IN AFRICA

DENIS GIKUNDA, LOCALIZATION PRG MANAGER

w3c: The Multilingual Web: Where are we?

Google in Africa
Local language content
Tools
Methodology (x 3)
GOOGLE IN AFRICA
WHAT, WHO, WHERE

• Making the internet an integral part of every-day life in Africa

• Access, Relevance, Sustainability

• Product Development, Engineering, Localization, Business Development, Marketing, PR, Sales*.  

+San-francisco, Zurich, London, New York, Dublin, Tel Aviv, Haifa
AFRICAN LANGUAGES

**landscape**
- Highest language density in world [2k+ languages]
- Over 100 languages with over 1M+ speakers
- 12 - 15 macro languages reach ~60% of indigenous language speakers
- Most use latin script, extended diacritics, with exception of Amharic (ET).

**Policy**
- English/French/Portuguese predominantly used as official or language of instruction in education
- Exceptions are Amharic (ET), Swahili (TZ), Setswana (BW), and 11 South African local languages.
- Large policy formulation gaps wrt language/education/ict, hence low demand for local language services. Potential partners are UNESCO, ANLOC, IDRC

**Status**
- African languages have remained a largely oral, informal phenomena. Very few books, newspapers, publications have been developed due to cost.
- Oral literature, indigenous knowledge, cultural novelty, and creativity remain unamplified, and lost over generations.
- Internet presents an opportunity to bootstrap written form of African languages.
Negligible african language content relative to speakers online

Stunted organic growth of content relative to user growth

Some efforts show promise of impact

<table>
<thead>
<tr>
<th>Year</th>
<th>New articles per day</th>
<th>Internet user growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Am</td>
<td>Sw</td>
</tr>
<tr>
<td>2006</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>2007</td>
<td>20</td>
<td>29</td>
</tr>
<tr>
<td>2008</td>
<td>75</td>
<td>29</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>2010</td>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>New articles per day</th>
<th>2000-2009</th>
<th>2000-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Am</td>
<td>2</td>
<td>13%</td>
<td>22%</td>
</tr>
<tr>
<td>Sw</td>
<td>29</td>
<td>42%</td>
<td>106%</td>
</tr>
<tr>
<td>Ar</td>
<td>61</td>
<td>165%</td>
<td>143%</td>
</tr>
<tr>
<td>Ru</td>
<td>529</td>
<td>239%</td>
<td>220%</td>
</tr>
<tr>
<td>Zh</td>
<td>185</td>
<td>246%</td>
<td>213%</td>
</tr>
<tr>
<td>En</td>
<td>1351</td>
<td>124%</td>
<td>110%</td>
</tr>
<tr>
<td>all langs</td>
<td>8457</td>
<td>226%</td>
<td>202%</td>
</tr>
</tbody>
</table>
• Users first generate content, or content that draws in users?
Automatic translation between 2,500+ language pairs
• Human translation between 100,000+ language pairs
• WYSIWYG display for MediaWiki text (not just Wikipedia)
• Direct publish to Wikipedia (preview mode only)

Google Sponsored Projects
Indic languages: 10MM+ words
Arabic: 5MM+ words
Swahili: 1MM+ words
AFC Leopards

From Wikipedia, the free encyclopedia

**AFC Leopards SC**, often simply known as "AFC" or "the Leopards," is a soccer club, based in Nairobi which was founded in 1964. Nicknamed "Ingwe" ("leopard" in Luhya), it is a traditional and popular Kenyan team and its home stadium is the Nyayo National Stadium, with a capacity of 30,000. Despite being based in the capital, Nairobi in the country's south east, the bedrock of its support is in western Kenya. The team plays in white shirts with blue hoops, white shorts and blue socks and sometimes with blue and white vertically-striped shirts; its alternative outfit is red jerseys and white shorts. Its traditional rival is Gor Mahia; their encounters are assumed to be thrillers, and because of the passion of the fans they are considered as the most intense and famous derbies in the region. AFC is considered an eastern, central and southern Africa soccer powerhouse because of its numerous continental appearances and its dominance in the CECAFA club championships, in addition to its dazzling array of formidable players.

### Contents [hide]

1. History
2. Support
3. Achievements
4. Performance in CAF competitions
5. Current squad
6. Club Officials
   6.1 Executive Committee
   6.2 Technical staff
7. References
8. External links

**AFC Leopards SC**

- **Full name**: AFC Leopards SC
- **Nickname(s)**: Ingwe
- **Founded**: 1964
- **Ground**: Nyayo Stadium, Nairobi, Kenya (Capacity: 30,000)
- **Chairman**: Julius Sikolo Ochieng
- **Coach**: Robert Matano
- **League**: Kenya Premier League

**Home colours**

---

**Away colours**

---
Support

AFC Leopards SC has a large and diverse fanbase, who hold a long-standing rivalry with other clubs, especially Gor Mahia. The club is arguably the biggest and most popular football club in Kenya and the Eastern Africa - with 9 million estimated supporters locally. These fans are composed into a number of supporters branches spread all over the country. More so, its support and fame transcends beyond the Kenyan boundaries to the rest of the continent especially the entire Eastern and Southern Africa. In a sense, Kenya and East Africa's favourite soccer club never plays an away match – AFC Leopards SC draws more supporters at away games than its opposition 'home' teams. Even at present the club still retains higher average attendances than all the clubs in the Kenya Premier League. In fact their first match upon return to the 2009 premier league had to be stopped because of insecurity occasioned by the collapsing of the perimeter fence due to a mammoth crowd of its supporters in and outside the ground.

Achievements

- **Kenyan Premier League**: 12
- **Kenyan President’s Cup**: 7
- **CECAFA Clubs Cup**: 5

Performance in CAF competitions

- **CAF Champions League**: 1 appearance
  - 1999 - First Round
- **African Cup of Champions Clubs**: 12 appearances
  - 1968: Semi-Finals
  - 1971: First Round

- **Ligi kuu ya Kenya**
- **Kombe la Rais**
- **Kombe la klubu za CECAFA:5**

Utendaji katika mashindano ya CAF.

- **CAF Champions League**: 1 kuonekana
  - 1999 - Raundi ya Kwanza
- **Kombe la Mbingwa wa klubu la Afrika**: Matookeo 12
AFC Leopards

AFC Leopards SC, mara nyingi hujulikana kama "AFC" au "the Leopards," ni klubu ya soka, mjini Nairobi, ambayo ilianzishwa mwaka wa 1964. Jina lake lingine ni "Ingwe" ("chui" katika lugha ya Kilithi), ni klubu maarufu na ya kitamaduni ya Kenya na uwanja wake wa nyumbani ni Uwanja wa Taifa wa Nyayo, ambayo ina nafasi ya watu 30000. Licha ya kuwa na makao katika mi mkuu, Nairobi katika sehemu ya kusini mashariki, ushabiki wake kwa wingi hutoka mkoa wa magharibi. Timu huvalia jesi nyeupe, vinyasa vyeupe na soksi za bluu na wakati mwingine na shati yenye rangi za bluu na nyeupe; nguo mbadala ni jesi nyekundu na vinyasa vyeupe. Mpinzani wake wa jadi ni Gör Mahia; makabiliano yao hufurahisha sana, na kutokana na mapenzi ya mashabiki makabiliano ya hawa wawili huchukuliwa kama "derby" maarufu na makali zaidi katika kanda. AFC inachukuliwa kama timu hatari ya Afrika ya mashariki, ya kati na ya kusini kwa sababu ya mechi nyingi ambazo imecheza barani na utawala wake katika michukani ya klubu ya CECAFA, kwa kuongeza wachegaji wake wazuri.

Yaliyomo [ficha]
1 Historia
2 Ufuasi
3 Majilio
4 Utendaji katika mashindano ya CAF
5 Mareje
6 Viungo vya nje

Historia

AFC Leopards ni mmojwapo wa vilabu ambayo vimefanikio zaidi katika kanda na hii ni ukweli kutokana mataji 12 ya ligi ambayo klubu hii imeshinda na kombe la Kenya mara 6 pamoja na shindano la klubu la CECAFA mara 5. Klabu hii ilianzishwa mwaka wa 1964 kama Abaluhya United FC wakati idadi ndogo ya vilabu viliumana Jina lililathishwa kuwa Abaluhya FC na baadaye kuwa All Footballers Federation-AFC, bado inadumisha AFC (na ikawa AFC leopards) mwaka wa 1980, wakati serikali ya Kenya ilipiga marufuku majina ya iliyopata vina milemela na mke na mwingine katika VIP channel na Greyhound TV na Ericsson.
COMMUNITY TRANSLATION

- **In a nutshell**
  - **Google Web Search Interface** in top 100 African languages.
  - **Translation Party model** - a fun, collaborative & social 2 day workshop involving students studying CS & language.

- **Use a toolkit** that combines MT, Glossary matching & global TM, and allows online collaborative work.
- **Quality is vetted** local language specialists, journalists, publishers.

- **Outcomes**
  - 300+ volunteers, 10+ Universities
  - 24 languages UIs launched.
  - Surge in search queries

- **Challenges**
  - Locale selection & disambiguation
  - Incentive / Reward
  - Glossary development
  - Internet Access

- **Approach**
  - Prioritize against internet penetration, usage status, content available. Inheritance, blind test,
  - Short term: Certificate, Training, Social, curriculum centered.
  - Long term: recognition, paid work.
  - Terminology harmonization, and release.

- **Outcomes**
  - 300+ volunteers, 10+ Universities
  - 24 languages UIs launched.
  - Surge in search queries
As the internet expands into low-penetration regions, demand for local language services & content grows.
In a nutshell

- **Wikipedia**: #3 content property globally (Alexa). 60% referrals from Google.
- **Contest**: grow Swahili Wikipedia articles by 500K words. Translate/author preselected, high traffic, substantive, relevant articles, using Google Translate/Google Translator Toolkit.
- **Partners**: 7 Universities in Kenya, Tanzania over 6 Week duration.
- **Prizes**: Netbooks, Internet modems, phones, and Google Schwag.

Challenges

- Process: Quality review, reversions, line by line translation.
- Technical: Published MT, markup,
- Sustained contribution
- References become multilingual?

Outcomes

- **Sw wiki pages**: 3/10 - 9/10
- +1600 Articles (+14%) | 7000 Articles in 10 months
- 1.9M words (100% CAGR), 800 registrants | 10 active contributors

Approach

- Content structure part of quality metric. Online training, using videos.
- MT as an enabler; prevent publishing with <50% human translation.
- Contest model. Partnership with decentralized Wikipedia Communities. Content focus (entertainment, local knowledge, sports)
In a nutshell

**Background**: High quality health information is particularly scarce in foreign languages, affecting arguably the most needy users.

**Volunteer effort** driven by Google.org. Participants are mainly medical student/faculty communities. Google matches every word in $1 of funding towards local health organization.

Targeting **Hindi**, **Arabic**, **Swahili** users

Outcomes

- **Approach**
  - Seed with paid translations, and professionally developed terminology to maximize TM leveraging in Google Translator toolkit.
  - Find partners with vested interest in the content.
  - Continue to work closely with decentralized communities -> Submit to talk page.

- **Challenge**
  - Audience/expertise disparity
  - Overwrites
  - Sustained Contribution

- **Outcomes**
  - ~1000 articles claimed
  - <10% published
  - >22,000 page views
  - >2000 registrants

"sitescontent.google.com/healthspeaks"
WHERE ARE WE?

Community
The community needs to be center stage for content to happen organically. Content will grow around communities needs.

Incentive / reward mechanisms
Should vary based on audience, content type and short/long term.
Short term: Contest prizes, accreditation, social networking. Longer term: Job opportunities, paid translation work.

Access
The cost of reliable PC based internet access is a real inhibitor to access. Will mobile be an enabler?

Tools / Platforms / APIs
Terminology & TM sharing via tools lower barrier for translation, allow more to participate.

Standards
Still lacking for African language wrt (i) variant/dialect classification (ii) term harmonization
• Discussion

• dgikunda@google.com

• @kariithi