

COMMUNITY TRANSLATION IN AFRICA

DENIS GIKUNDA, LOCALIZATION PRG MANAGER w3c:The Multilingual Web: Where are we?

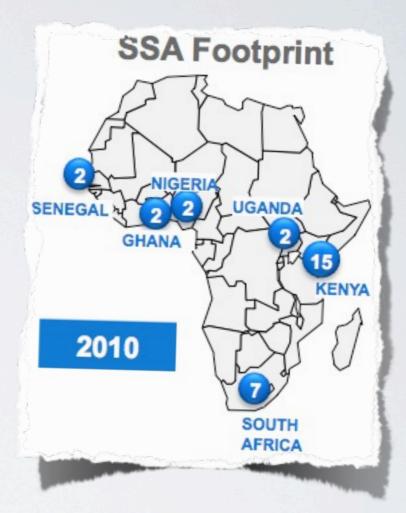
Google in Africa
Local language content
Tools
Methodology (x 3)



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).

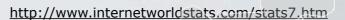
Policy

- 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 a opportunity to bootstrap written form of african languages.





http://stats.wikimedia.org/EN/



Stunted organic growth of content relative to user growth

Some efforts show promise of impact



Native speakers online (M)

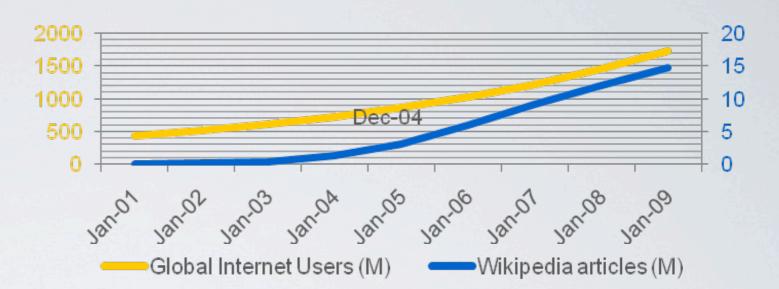
Wikipedia articles (K)

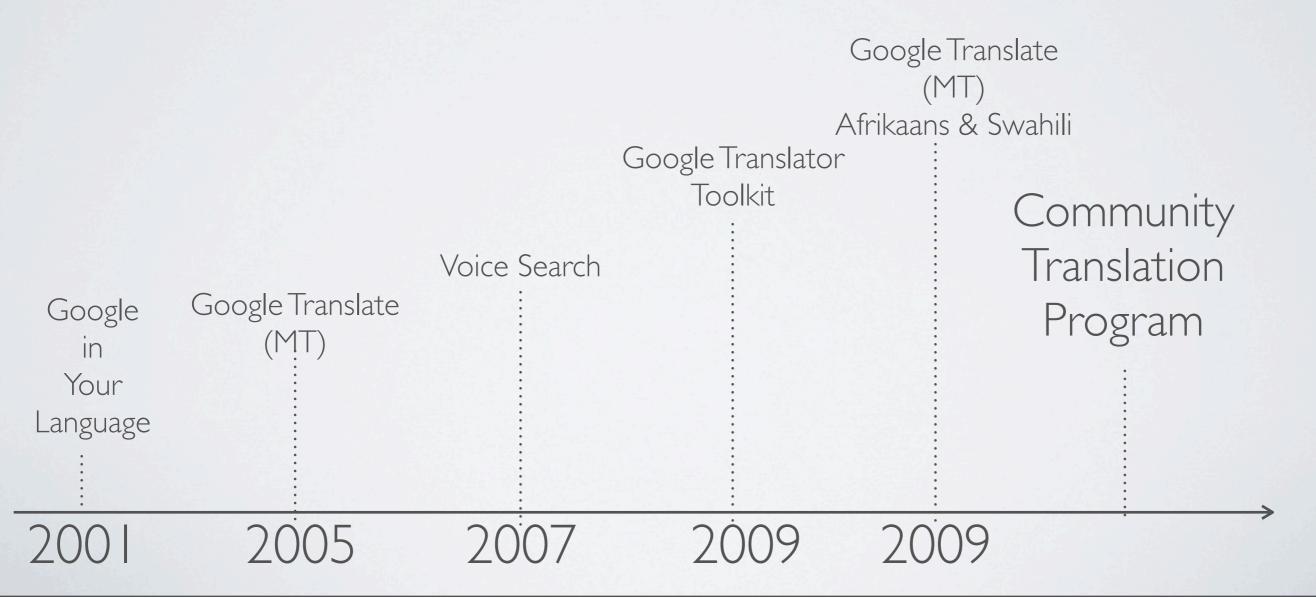
	New articles per day	Internet user growth	2000-2009	2000-2010
am	2	2810%	13%	22%
sw	29	247.8%	42%	106%
ar	61	1545%	165%	143%
ru	529	1125.8%	239%	220%
zh	185	894.8%	246%	213%
en	1351	226.7%	124%	110%
all langs	8457	342.2%	226%	202%

USER GENERATED CONTENT



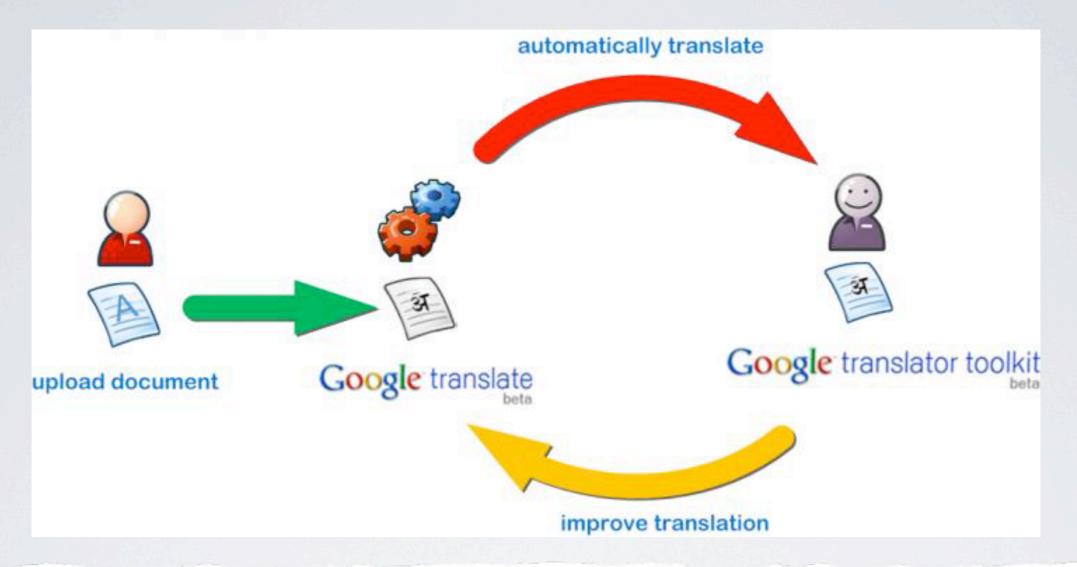
 Users first generate content, or content that draws in users?







TOOLS



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: IOMM+ words

Arabic: 5MM+ words

Swahili: IMM+ words



Google confidential & proprietary



Main page Contents Featured content Current events Random article Donate

- Interaction About Wikipedia Community portal Recent changes Contact Wikipedia Help
- ▶ Toolbox
- Print/export
- Languages Català Français Kiswahili Nederlands

Article Discussion

Read Edit View history

Search

Q

AFC Leopards

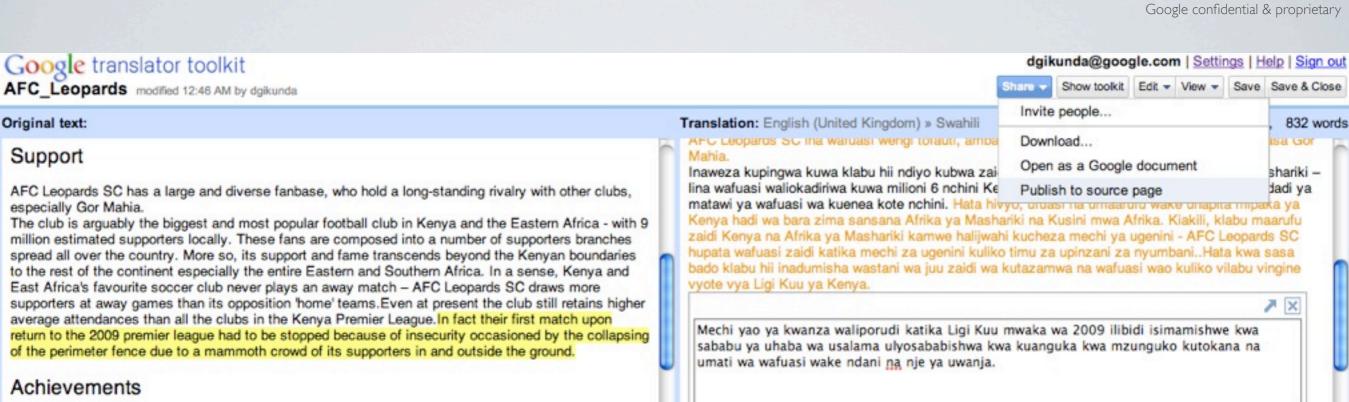
From Wikipedia, the free encyclopedia

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.





«Previous Next»



Kenyan Premier League: 12

1966, 1967, 1970, 1973, 1980, 1981, 1982, 1986, 1988, 1989, 1992, 1998

Kenyan President's Cup: 7

1967, 1968, 1985, 1991, 1994, 2001, 2009

CECAFA Clubs Cup: 5

1979, 1982, 1983, 1984, 1997

Performance in CAF competitions

CAF Champions League: 1 appearance

1999 - First Round

African Cup of Champions Clubs: 12 appearances

[5]

1968: Semi-Finals 1971: First Round

Majalio

Ligi kuu ya Kenya

1966, 1967, 1970, 1973, 1980, 1981, 1982, 1986, 1988, 1989, 1992, 1998

Kombe la Rais

1967, 1968, 1985, 1991, 1994, 2001, 2009

Kombe la klabu za CECAFA:5

1979, 1982, 1983, 1984, 1997

Utendaji katika mashindano ya CAF.

CAF Champions League: 1 kuonekana

1999 - Raundi ya Kwanza

Kombe la Mabingwa wa klabu la Afrika : Matokeo 12

Q

Google confidential & proprietary



Mwanzo Jumuia Matukio ya hivi karibuni Mabadiliko ya karibuni Ukurasa wa bahati Msaada Michango

- Viungo viungavyo ukurasa huu Mabadiliko husika Kurasa maalum Ukurasa wa kuchapika Kiungo cha daima Cite this page
- Lugha zingine Català English Français Nederlands

Makala Majadiliano

Soma Hariri Fungua historia

Kutafuta

AFC Leopards

AFC Leopards SC, mara nyingi hujulikana kama "AFC" au "the Leopards," ni klabu ya soka, mjini Nairobi, ambayo ilianzishwa mwaka wa 1964. Jina lake lingine ni "ingwe" ("chui" katika lugha ya Kiluhya), ni klabu 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 mji 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 ya 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 michuano ya klabu ya CECAFA, kwa kuongezea wachezaji wake wazuri.

Logo Full jina AFC Leopards SC Nickname(s) Ingwe Kimeanzishwa 1964 Ground Nyayo Stadium Nairobi, Kenya (Capacity: 30,000) Cheyaman Alex ole Magelo Ligi Kenya Premier League Home colours Away colours

AFC Leopards SC

Yaliyomo [ficha]

- 1 Historia
- 2 Ufuasi
- 3 Majilio
- 4 Utendaji katika mashindano ya CAF
- 5 Marejeo
- 6 Viungo vya nje

Historia [hariri]

AFC Leopards ni mmojwapo wa vilabu ambavyo vimefanikio zaidi katika kanda na hii ni ukweli kutokana mataji 12 ya ligi ambayo klabu hii imeshinda na kombe la Kenya mara 6 pamoja na shindano la klabu la CECAFA mara 5. Klabu hii ilianzishwa mwaka wa 1964 kama Abaluhya United FC wakati idadi ndogo ya vilabu viliungana Jina lilibadilishwa 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

Google confidential & proprietary

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, **IO+** Universities

24 languages Uls launched.

Surge in search queries



Challenges

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

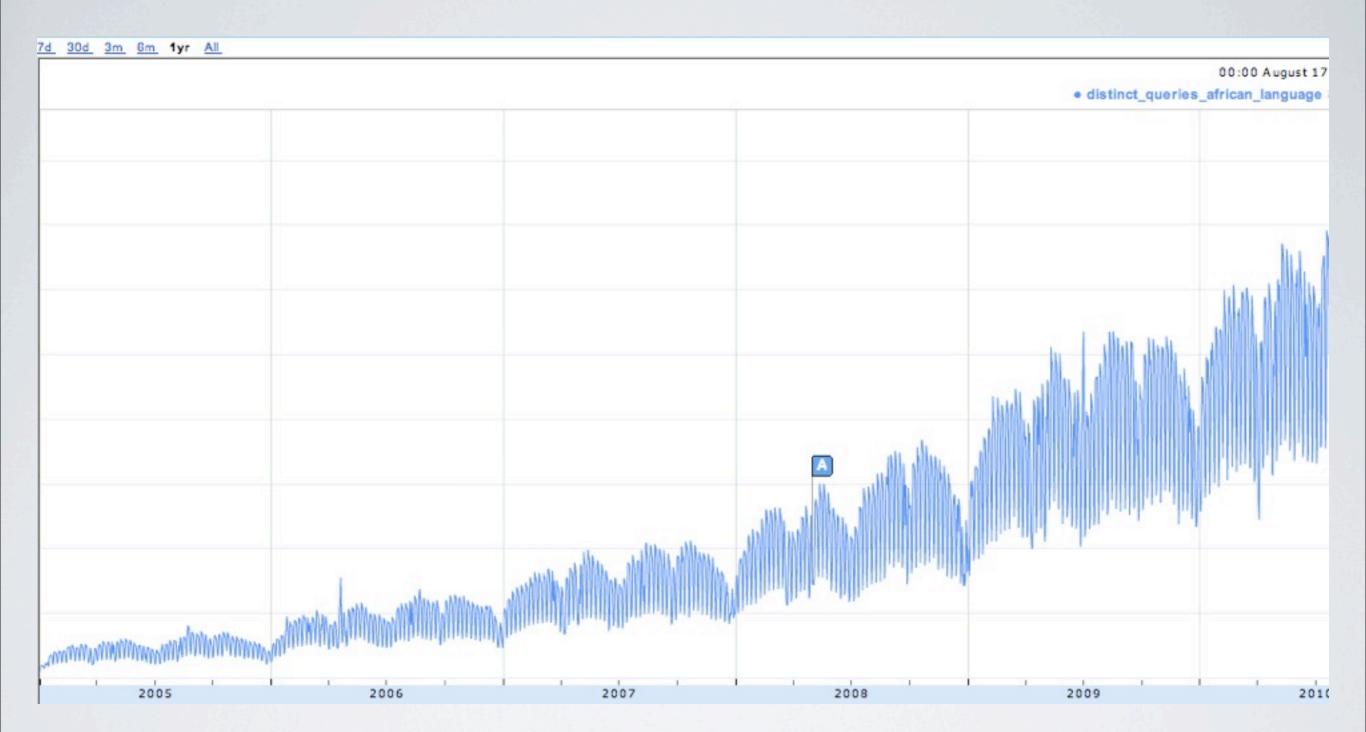
Approach

- 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.







A - SSA community Translation program begins
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



+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)

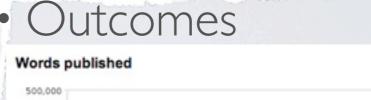


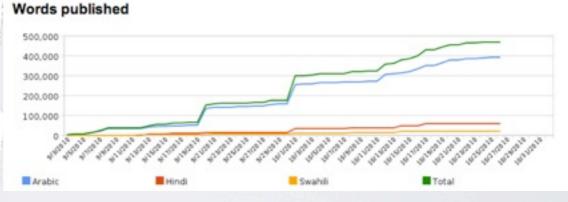
HEALTHSPEAKS

Afya 健康 Zdravje

sitescontent.google.com/healthspeaks

- 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
- Challenges
- Audience/expertise disparity
- Overwrites
- Sustained Contribution





- ~1000 articles claimed <10% published
- >22,000 page views

>2000 registrants

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.



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



Google confidential & proprietary

- Discussion
- dgikunda@google.com
- @kariithi



