



ADAPTING EMM TO SWAHILI

SWAHILI

The name 'Kiswahili' comes from the plural *sawāḥil* (سواحل) of the Arabic word *ṣāḥil* (ساحل), meaning "boundary" or "coast".

Bantu language with borrowed words from Arabic, German, Portuguese and English.

A type of language where the grammar arranges nouns according to classes (such as people, tools, animals).

Spoken by over 50 million people in sub-Saharan regions.



SWAHILI spoken in:

BURUNDI

CONGO

KENYA

MOZAMBIQUE

RWANDA

SOMALIA

TANZANIA

Official language in:

KENYA

TANZANIA

UGANDA

KISWAHILI

MOTIVATION

- Collaboration with African Union
- Collaboration with Pan African Parliament
- To extend EMM to more languages

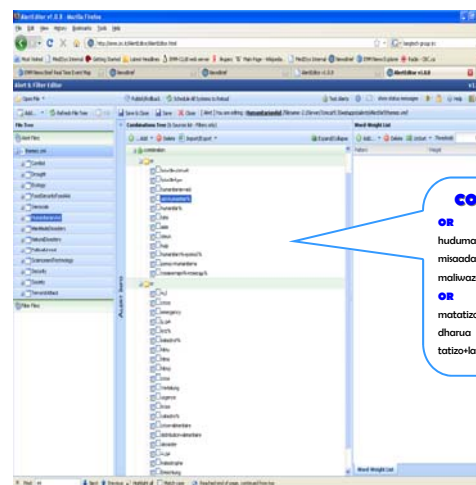
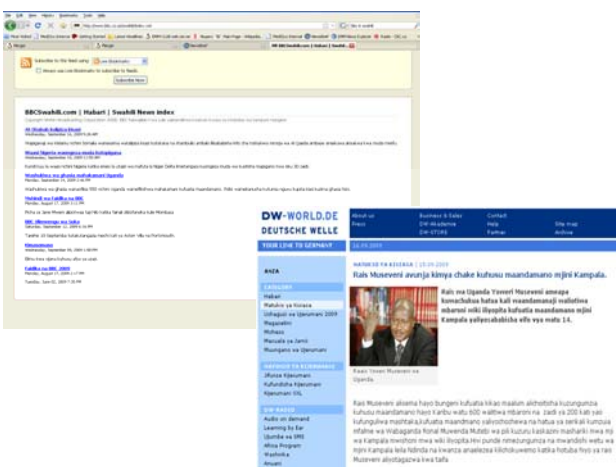


HOW TO PROCEED?

- Find the right news sources and integrate them in the system
- Update the category definitions-Boolean operators with optional proximity operator and wild cards (countries, EMM themes...)
- Update the entity recognition resources (common first names, professions...)
- Adopting the EMM user interface to Swahili

CHALLENGES

- Some news sources are not in text format (html/rss)
- Scarcity of news sources
- Language features (noun classes, no affixes on proper names, affixes on verbs)
- Variation of Swahili from one country to another (dialects)



COMBINATION

- OR
- huduma muhimu
- misaa
- malwazo
- OR
- matatizo
- dharau
- tatizo la chakula

CONTRIBUTORS

Monica De Paola, Janya Belyayeva, Sylvia Ombuya, OPTIMA - European Commission - Joint Research Centre - Institute for the Protection and the Security of the Citizen

E-mail Format: Firstname.Lastname@ec.europa.eu

