

## OPEN SOURCE KISWAHILI SPELL CHECKER (SW-TZ)

**Hashim M. Twaakyondo<sup>1</sup> and Kennedy Mwakisole<sup>2</sup>**  
Centre for Virtual Learning, University of Dar es Salaam  
P.O.Box 35062, Dar es Salaam, Tanzania  
e-mail: <sup>1</sup>[htwaaky@udsm.ac.tz](mailto:htwaaky@udsm.ac.tz), <sup>2</sup>[kenfra@udsm.ac.tz](mailto:kenfra@udsm.ac.tz)

### ABSTRACT

*Different English software products are localized into many native languages spoken around the world, the most popular software products localized so far are word processing software and web browsers. The effort has begun to localize these software products in Kiswahili language which is widely spoken in Tanzania, East and Central Africa. Kiswahili is an official language of Tanzania, and is a national language in Uganda and Kenya. To make these software products useful to Kiswahili speaking community particularly in Tanzania a new Kiswahili spell checker has been developed and, it uses most of the Kiswahili words spoken in Tanzania and its locale identity is sw-TZ for OpenOffice.org and Mozilla Firefox while Kiswahili spoken in Kenya has a locale identity sw-KE which is common for Microsoft office products. It was developed due to the shortcomings of first ever Kiswahili spell checker developed by Open Kiswahili Localization Project (KILINUX). A spell-checking tool called MySpell has been used in developing a new spell checker. The dictionary of a new spell checker consists of 4894 stems and its affixes were created based on Kiswahili noun classification system and by selecting different groups of verb stems that follow the same trends in generating other Kiswahili valid words. This paper addresses the shortcomings of the existing spell checker and the approach carried out to develop a new spell checker.*

**Keywords:** Kiswahili, MySpell, OpenOffice, Mozilla Firefox, Kilinux

### INTRODUCTION

Kiswahili is a Bantu language, generally spoken by more than 130 million people all over the world. It is both official and national language of Tanzania and Kenya (Sewangi, 2006). Students in Tanzania particularly in primary schools are taught by using Kiswahili as the medium of instruction and Kiswahili is taught as a lesson in secondary schools. At the universities, for instance University of Dar es Salaam offers degree programmes in Kiswahili (IKR, 1930). In government institutions and organizations such as parliament, Kiswahili is used as the official language. The most useful software in both government and public sectors are office suites and web browsers

and the well known offices are Openoffice.org and Microsoft office while the web browsers are Mozilla Firefox and Internet Explorer.

The efforts are underway to localize these software products into Kiswahili (KILINUX, 2008). The localized software needs to be integrated with developed Kiswahili spell checker that will enable Kiswahili users to spell words while they are typing e-mails and documents in Kiswahili and doing their work faster.

#### **Kiswahili language**

Kiswahili, like other Bantu languages, has an extensive system of prefixes, suffixes and infixes, particularly found on verbs and to some extent derived nouns and

adjectives. Most of these affixes are grammatically obligatory although some can be optional in some circumstances, in some cases depending on stylistic variation. Kiswahili also has a richly inflected noun class system. Prefixes and infixes must agree with the relevant noun, whether they are on adjectives, verbs, or pronouns. As with many other Bantu languages, Kiswahili has only five distinct vowels (a,e,i,o,u), and it is therefore possible to write the language using the Roman alphabet.

Kiswahili has 23 consonants (b, c, d, f, g, h, i, j, k, l, m, n, p, r, s, t, v, w, y, th, dh, gh, sh) and the nasal consonants in Kiswahili are (ny, ng', ng, nd, mb, mw). Only Kiswahili noun classification and verbs are briefly explained below. Table 1 shows different classes of nouns used in Kiswahili language and they are classified according to their noun prefixes, subject prefixes and object prefixes.

**Table 1:** Kiswahili Noun Class System

Class	Noun Prefix	Subject Prefix	Object Prefix	Examples				
1	m-/mw-	a-; yu-	m-/mw-	Mkubwa	mpya	Mweusi	mwingine	mwenye
2	wa-/w-	wa-/w-	wa-	Wakubwa	wapya	Weusi	wengine	wenye
3	m-/mw-	u-/w-	u-	Mkubwa	mpya	Mweusi	mwingine	wenye
4	mi-/mw-	i-/u-	i-	Mikubwa	mipya	Myeusi	mingine	yenye
5	th,ji-/j-	li-/l-	li-	Kubwa	jipya	Jeusi	jingine lingine	lenye
6	ma-/m-	ya-/y-	ya-	Makubwa	mapya	Meusi	mengine	yenye
7	ki-/ch-	ki-/ch-	ki-	Kikubwa	kipya	Cheusi	kingine	chenye
8	vi-/vy-	vi-/vy-	vi-	Vikubwa	vipya	Vyeusi	vingine	vyenye
9	n-/m-/th-	i-/y-	i-	kubwa nzuri	mpya	Nyeusi	nyingine	yenye
10	n-/m-/th-	zi-/z-	zi-	kubwa nzuri	mpya	Nyeusi	nyingine	zenye
11	u-/w-	u-/w-	u-	mkubwa	mpya	Mweusi	wingine	wenye
14	u-/w-	u-/w-	u-	mkubwa ukubwa	mpya upya	mweusi weusi	mwingine	wenye
15	ku-/kw-	ku-/kw-	ku-	kukubwa	kupya	Kweusi	kwingine	kwenye
16		pa-/p-	pa-	pakubwa	papya	Peusi	pengine	penye
17		ku-/kw-	ku-				kwingine	kwenye
18		mu-/m-	mu-				mwingine	mwenye

### Kiswahili Verb

The Kiswahili verb consists of a subject marker, a tense marker and a verb stem.

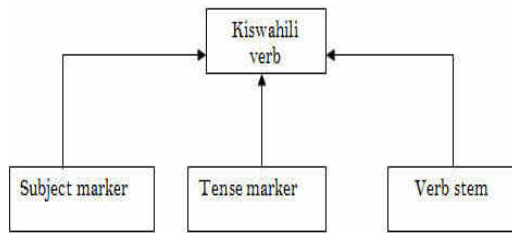


Figure 1: Structure of Kiswahili Verb

For instance, the word "Ninalala" can be split up into three parts fitting the above category.

Ni-na-lala I am sleeping  
Ni - is the subject marker for "I"  
na - is the tense marker indicating "present tense"  
lala - is the verb stem for "sleep"

The verb in Kiswahili is marked by a prefix which identifies the subject. The following are some of the various prefixes.

ni - I - First person singular  
u - you - Second person singular  
a - she/he - Third person singular  
tu - we - First person plural  
m - you - Second person plural  
wa - they - Third person plural

In referring to the "present tense" the tense marker "na" is used

Thus: Ninalala, unalala, analala, tunalala, mnalala, wanalala

In referring to the "past tense" the tense marker "li" is used.

Thus: nililala, ulilala, alilala, tulilala, mlilala, walilala.

The future tense marker is "ta"  
e.g., nitalala, tutalala, etc.

### Open Source Software (OSS)

Open-Source software is the software in which the source code is distributed or accessible via the internet without charge or limitations on modification and future distribution by third parties (Christian *et al.*, 2004), the most known open source

software are office suite called Openoffice.org (OPENOFFICE.ORG, 2003) and web browser called Mozilla Firefox (MOZILLA FIREFOX, 1998).

The second paradigm of software development concept is proprietary software development. Unlike open-source software, it is paid for and does not come with the source code. Proprietary companies that develop such software employ software developers for building and fix their software in a closed environment. The popular closed source software are office suite called Microsoft Office and web browser called Internet Explorer. The main feature that characterized open source software is the freedom users have, that is:

- The ability to customize software to local languages and cultures
- Lowered barriers to entry for software businesses
- Participation in global network of open source software development.

These are the good advantages to the Kiswahili community, because Kiswahili is the widely spoken language.

### Jambo OpenOffice.org 1.1.3

Jambo OpenOffice.org 1.1.3 is the first release of an Office Suite in Kiswahili. The software has been developed/localized based on the free and open source software OpenOffice.org 1.1.3. (KILINUX, 2008).

### Jambo Mozilla Firefox 1.0.3

Jambo Mozilla Firefox 1.0.3 is the first release of Kiswahili web browser. The software has been developed/localized based on the free and open source web browser called Mozilla Firefox 1.0.3. (KILINUX, 2004).

The two Kiswahili software products were developed/localized by Open Kiswahili Localization Project (Kilinux). The following are ten phases in summary used by Kilinux Project towards the localization process of English software

products into Kiswahili (Twaakyondo and Escudero, 2004).

- ◆ Selection of the Software to be localized.
- ◆ Extraction and Conversion of Strings.
- ◆ Translation of the Strings.
- ◆ Reviewing the Translated Strings.
- ◆ Insertion of the Translated Strings into the code.
- ◆ Creation of new localized code (first release of localized code is built).
- ◆ Reviewing the localized code.
- ◆ Creation of new localized code (final release of localized code is built).
- ◆ Publication of the Localization.
- ◆ Maintenance of the localized code.

**Related work**

The existing Kiswahili spell checker was released in December 2004. The spell checker consists of 67900 Kiswahili words contributed by different Kiswahili experts (JAMBO SPELL CHECKER, 2004). It was developed using MySpell, this tool allows stems in the wordlist to combine with suffixes/prefixes defined in the affix file (LINGUCOMPONENT, 2003). Figure 2 shows the portion of a wordlist used by the existing Kiswahili spell checker.

67900	Aliitaja	Bonde
Ababa	Allah	Buki
Abdallah	Azori	Buleani
Abiramu	Baali	Bungoma
Abiudi	Babuloni	Bute
Adamu	Babulonia	kataa
Adi	Badar	waliwasimami
Aha	Bahari	waliwaunga
Afikir	Bajini	waliwavuruga
Afya	Bernike	waliyokataa
Ahero	Biblia	waliyofundis
Akichangia	Bidaya	unga
Alidhaniwa	Blasto	

**Figure 2:** The portion of the wordlist used in the existing spell checker.

The affix file of the existing developed Kiswahili spell checker consisted of only four prefixes as shown in Figure 3.

PFX A Y 4
PFX A 0 wali .
PFX A 0 ali .
PFX A 0 tuli .
PFX A 0 zili .

**Figure 3:** The portion of the affix file that shows four prefixes

**Shortcomings of the first developed Kiswahili spell checker Wordlist**

Figure 3 shows some of the words which are in the wordlist of the existing spell checker. The wordlist consists of both stems and the inflected/derived words made when stems allowed to combine with prefixes or suffixes defined in the affix file. For example, “kataa” and “unga” were the correct stems to be in a wordlist but a word “waliyokataa” was made when a stem “kataa” was combined with a prefix “waliyo” while a word “waliwaunga” was made when a prefix “waliwa” was combined with a stem “unga”. The two words “waliyokataa” and “waliwaunga” were not supposed to be in a wordlist as it is in the existing Kiswahili spell checker; instead the spell checker should generate them automatically during the process of analyzing and generating suggestions corresponding to misspelled words. Also the wordlist has misspelled or incomplete words, for example, the word “waliyofundis” should be “waliyofundisha” and “waliwasimami” should be “waliwasimamia” (JAMBO SPELL CHECKER, 2004). Moreover, the wordlist contains names of people such as “Chiligati” and “Ishengoma”, which were not supposed to be in it, because only selected and well known names of people, cities and town are allowed to be in the wordlist.

The existing Kiswahili spell checker consists of an affix file that defined only few prefixes and did not include Kiswahili suffixes as shown in Figure 2. The spell checker does not use this file because the defined prefixes were already

combined with stems to form other words that were also added to the wordlist. Therefore, there is no automatic combination of stems in the wordlist and prefixes defined in the affix file during the process of analyzing and generating suggestions corresponding to the misspelled Kiswahili words.

According to the wordlist and affix file of the existing Kiswahili spell checker, the inflected Kiswahili words from a Kiswahili stem “kula” were “alikula”, “walikula”, “tulikula” and “zilikula”. This means that the derived words follow the following structure.

“a” is a subject marker for third person singular.

“wa” is a subject marker for third person plural.

“tu” is a subject marker for first person plural.

“zi” is a subject marker for third person plural.

“li” past tense marker.

“kula” is a verb stem.

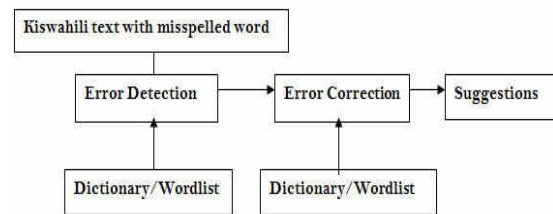
These four prefixes used only three subject markers, other three subject markers which are “u” for second person singular, “ni” for first person singular and “m” second person plural were excluded. Moreover, these prefixes created only past tenses and left behind other tenses, for example present and future tenses. Inflected words were also added to the wordlist. Therefore, a spell checker relies only on a wordlist for error verification and correction.

The structure of the existing Kiswahili spell checker might be as follows.

In Figure 4, a spell checker uses only a dictionary/wordlist in the whole process of error detection for Kiswahili text with misspelled words and uses the same wordlist in generating suggestions.

**Table 2:** Inflected words in the existing Spell Checker

Subject Marker	Past Tense Marker	Verb Stem
a	li	kula
wa	li	kula
tu	li	kula
zi	li	kula

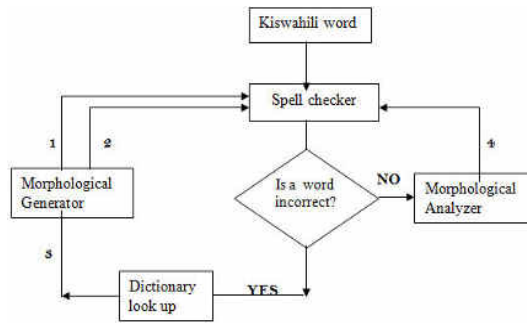


**Figure 4:** Structure of existing Spell Checker

**DEVELOPMENT OF THE NEW KISWAHILI SPELL CHECKER**

A new Kiswahili spell checker was developed using the same spell-checking tool for Openoffice.org, this spell checker consists of both the wordlist and affix file. It allows stems stored in wordlist to combine with prefixes and suffixes defined in the affix file during the process of analyzing and generating suggestions corresponding to the misspelled words (LINGUCOMPONENT, 2003).

In developing this Kiswahili spell checker, the approach used is as follows; Firstly, the Kiswahili wordlist (dictionary) that stores Kiswahili stems was developed and secondly, the affix file which stores Kiswahili prefixes/suffixes was developed. Figure 5 shows the structure of the new Kiswahili spell checker.



**Figure 5:** Structure of the new Spell Checker

**Key to Numbering in Figure 5.**

1. Suggestion list
2. Stem + affixes
3. Word with corrected affixes
4. Stem + affixes

***Kiswahili Dictionary (Wordlist) of the new Spell Checker***

The Kiswahili dictionary used in the new Kiswahili spell checker was made up of 4894 stems (KILINUX, 2008). These stems were extracted from the Kiswahili glossary which was in PO (Poedit file) format. The file has more than 4000 English terms translated into Kiswahili language. Table 3 shows a portion of the file from which stems were extracted.

The translated Kiswahili terms which appear next to “msgstr” part in the `kiswahili_glossary.po` file were extracted. The script which was used to extract the Kiswahili terms and arrange them in alphabetic order was developed using Perl programming language.

When the script was run against `kiswahili_glossary.po` file, a wordlist was generated with all terms arranged in alphabetic order. The file generated by the above script was named `sw_TZ.dic`. Table 4 shows portion of the `sw_TZ.dic` file generated by the Perl script.

Currently the dictionary file (wordlist) contains only verb stems and singular noun stems. Each term in the wordlist was checked to make sure that only (stems) remained in `sw_TZ.dic` file, which was

different from the existing Kiswahili spell checker in which stems and derived/inflected words were included in the wordlist.

**Table 3:** The Portion of Kiswahili ICT Glossary from Kilinux Project

<code>kiswahili_glossary.po</code>
"Project-Id-Version: \n"
"POT-Creation-Date: \n"
"PO-Revision-Date: 2005-10-10 09:59+0300\n"
"Last-Translator: Kilinux Team <kilinux_team@kilinux.org>\n"
"Language-Team: \n"
"MIME-Version: 1.0\n"
"Content-Type: text/plain; charset=iso-8859-1\n"
"Content-Transfer-Encoding: 8bit\n"
msgid "across"
msgstr "kukingama"
msgid "actual"
msgstr "halisi"
msgid "advance"
msgstr "endelea"
msgid "after"
msgstr "baada"
msgid "align"
msgstr "fungamana"
msgid "amount"
msgstr "kiasi"

**Table 4:** The Portion of Kiswahili Wordlist

4894	aya	bana	barakoa
athari	baada	bananga	barua
athiri	badili	bandia	batili
aula	bafa	bandika	baiti
avya	bahasa	banua	
awali	baiti	baobonye	

***Affix Creation***

An affix is either a prefix or a suffix attached to a stem to make other valid words. For example, in Kiswahili word “kula” a prefix might be “ana”. When it is attached to the stem “kula”, another Kiswahili word “anakula” will be generated. In another Kiswahili stem “kamili”, the suffix might be “fu” and when it is attached to the stem “kamili”, another Kiswahili word “kamilifu” will be generated.

**Table 5:** Prefix “wa”

Singular	Plural
mwana	Wana
mwalimu	walimu
mlinzi	walinzi
msomi	wasomi
muuguzi	wauguzi
mwizi	wezi

The process of creating Kiswahili morphological rules (affixes) was carried out on the basis of the knowledge acquired from the Kiswahili language structure and morphology. First, affixes were created based on Kiswahili noun classification system and secondly, affixes were created by selecting different groups of verb stems which follow the same trends in generating other Kiswahili valid words. The following sections explain in brief the creation of Kiswahili morphological rules (affixes).

**The creation of rules based on Kiswahili noun classification system**

In this case singular nouns were taken as noun stems and added directly to the wordlist while plural nouns were taken as inflected words from existing singular nouns. Explanations per noun classification are given in the following sections:

**Noun class 1-2**

For the noun class 1 and 2, all the nouns class 1 were taken as stems and added to the wordlist while nouns of class 2 were generated from class 1, i.e. noun class 1 were singular and noun class 2 were plural as shown in Table 5.

The rules are;

1. For any noun class 1, stem starts with “m” followed by any consonant except “w” then strip off “m” and add “wa”.
2. For any noun class 1, stem starts with “m” followed by either “w” or “u”

then strip off either “mw” or “mu” and add “wa”.

3. For any noun class 1, stem starts with “m” followed by “w” and either “e” or “i” then strip off “mwi” or “mwe” and add “we”.

**Noun class 3-4 (Names of tree and things that spread)**

For the noun class 3 and 4, all the nouns in class 3 were taken as stems and added to the wordlist while nouns in class 4 were generated from class 2, i.e., noun class 2 were singular and noun class 4 were plural.

**Table 6:** Prefix “mi”

Singular	Plural
mti	miti
mwanzi	mianzi
moto	mioto
msaada	misaada
mwiko	miiko

The rule is;

For any noun class 3, stem starts with “m” or “mw” then strip off “m” or “mw” and add “mi” as illustrated in table 6.

**Noun class 7-8**

For noun class 7 and 8, all the nouns in class 7 were taken as stems and added to the wordlist while nouns in class 8 were generated from class 7, i.e., noun class 7 were singular and noun class 8 were plural.

**Table 7:** Prefix “vi” and “vy”

Singular	Plural
kiti	viti
kikombe	vikombe
kifuniko	vifuniko
chakula	vyakula
chombo	vyombo
chandarua	vyandarua

The rules are;

1. For any noun class 7, stem starts with “ki” then strip off “ki” and add “vi”.

- For any noun class 7, stem starts with “ch” followed by vowel except “i” then strip off “ch” and “vy”.

Creation of rules by selecting different groups of verb stems which follow the same trends in generating other Kiswahili valid words.

**Table 8:** Suffix “o”

andika	andiko
panga	pango
ingiza	ingizo
chagua	chaguo
biringiza	biringizo

Table 8 shows few Kiswahili verb stems selected from a wordlist of the new Kiswahili spell checker. The simple rule based on the selected stems might be: For any verb stem in the wordlist that ends with an “a” then strip off the “a” and add an “o”.

**Table 9:** Suffix “iwa” and “wa”

panga	pangiwa
andika	andikiwa
pakia	Pakiwa
funga	fungiwa
unga	ungiwa
pungia	pungiwa

According to Table 9, the rule is: For a verb stem that ends with letter “i” and “a” then strip off the “a” and add a “wa”, while for any verb stem that does not end with any of the characters in [aeiou] and the “a”, then strip off the “a” and add an “iwa”.

Figure 6 shows the portion of Kiswahili affix file which contains some of the created affix rules that will be applied to the stems available in the Kiswahili wordlist. The affix file. The characters representing flags in the dictionary file are in uppercase.

```

SET ISO8859-1
TRY
aiunkemohwtlsgybzpdrfjcv'KMSAWTLBNE
YDUGHPFIROZJC-V

PFX M Y 5
PFX M m wa [aeiouw]
PFX M mu wa mu
PFX M mw wa mw
PFX M mwe we mwe
PFX M mwi wa mmwi

PFX C Y 2
PFX C ki vi vi
PFX C ch vy ch(aeou)

SFX W Y 2
SFX W a iwa [^aeiou]a
SFX W a wa [eiou]a

SFX S Y 3
SFX S a isha a
SFX S u isha u
    
```

**Figure 6:** The Portion of Kiswahili Affix file (sw\_TZ.aff)

**Specifying Affix Flags**

Affix flags are specified in the wordlist by specifying them after the “/” character i.e. *word/flags*. For example, amili/FS associates the “F” and “S” flags with the stem “amili”. Figure 7 shows the Kiswahili dictionary file which consists of both stems and some flags. Letters A, F, S and U are the names of the characters that represent either prefix or suffix defined in uppercase.

**RESULTS AND TESTING THE SPELL CHECKER**

The spell checker results were tested using the words which were in the Kiswahili wordlist/dictionary file. In this case a simple text file consisting of Kiswahili words was created and tested against the spell checker to check the misspelled words and generate suggestions. A sample text file is shown in Figure 8.



4894	alama	amuru	apuleti
acha/S	alamampaka	andaa	aridhia
adapta	alamisha	andika	arifu
aga	alfabeti	angaanga	asilimia
agoza	alika/U	angalia/A	athari
agumeni	algorithithi	angaza/AS	awali/F
ainisha	amali	anuani	badili/S
akaunti	amili/FS	anuwai	
akisi/A	ampasendi	anza/SUA	
akronomia	amri	apostrofi	

**Figure 7:** The portion of the new Kiswahili dictionary file with flags

achishe	angaliwe
amilife	angalia
aliki	angazo
mwanzi	angaz
amilishua	chakata
pangisha	mchakato
panguo	anzisha
pangui	amili
mchakati	kamilifu
	hamasisha

**Figure 8:** Maneno file

Figure 8 shows some Kiswahili words which were selected and tested against the new Kiswahili spell checker. After running the new spell checker against maneno text file the output was as shown below.

This is the *output.txt* file:

"achishe" neno haliko sawa!

mapendekezo:

... "achisha"  
... "achisho"  
... "achishwa"

"amilife" neno haliko sawa!

mapendekezo:

... "amilifu"  
... "amilisha"

"aliki" neno haliko sawa!

mapendekezo:

... "alika"  
... "alika"  
... "aliko"

"mwanzi" neno haliko sawa!

mapendekezo:

... "mwanza"  
... "mwanzo"

"amilishua" neno haliko sawa!

mapendekezo:

... "amilisha"

"pangisha" neno liko sawa

"panguo" neno haliko sawa!

mapendekezo:

... "pangua"  
... "pango"

"pangui" neno haliko sawa!

mapendekezo:

... "pangua"

"mchakati" neno haliko sawa!

mapendekezo:

... "mchakata"  
... "mchakato"

"angaliwe" neno haliko sawa!

mapendekezo:

... "angalia"  
... "angaliwa"  
... "kuangalia"

"angalia" neno liko sawa

"angazo" neno haliko sawa!

mapendekezo:

... "angaza"

"angaz" neno haliko sawa!

mapendekezo:

... "angaza"

"mchakato" neno liko sawa

"anzisha" neno liko sawa

As it appears in the *output.txt* file, the misspelled words were analyzed by the system which then generated the suggestions.

### Integration of the Spell Checker with Jambo OpenOffice.org 1.1.3

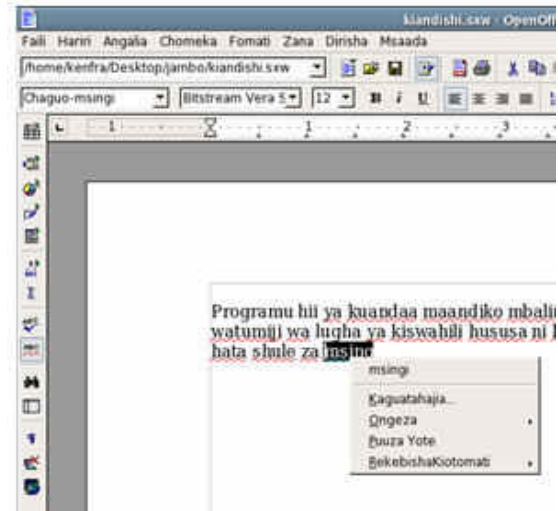
When Jambo OpenOffice.org was installed on a computer running Linux Operating system the following path `../OpenOffice.org 1.1.3/share/dict/ooo` was created automatically. Therefore to integrate the spell checker with Jambo OpenOffice both Kiswahili wordlist (`sw_TZ.dic`) and affix (`sw_TZ`) files were copied to the directory `../OpenOffice.org 1.1.3/share/dict/ooo` then Jambo OpenOffice restarted. To make sure that the spell checker was properly loaded, the two Kiswahili files were renamed as follows:

`sw_TZ.dic` was renamed to `sw_KE.dic`.  
`sw_TZ.aff` was renamed to `sw_KE.aff`.

This was done due to the fact that only Kiswahili language spoken in Kenya (`sw_KE`) was registered by OpenOffice.org and by default in the language selection parts only `sw_KE` is the naming standard used by OpenOffice.org when a user selects to use the Kiswahili version of OpenOffice.org. Kiswahili spoken in Tanzania (`sw-TZ`) is not registered as one of the languages used by OpenOffice.org and the efforts are underway to register it. In the Linux command prompt the following Linux commands were used for both copying and renaming Kiswahili wordlist and affix files:

```
# cp sw_TZ.aff /OpenOffice.org
1.1.3/share/dict/ooo
# cp sw_TZ.dic /OpenOffice.org
1.1.3/share/dict/ooo
#mv sw_TZ.aff sw_KE.aff
#mv sw_TZ.dic sw_KE.dic
```

Figure 9 shows how the new Kiswahili spell checker works with Jambo OpenOffice.org 1.1.3.



**Figure 9:** Kiswahili Spell Checker integrated with Jambo OpenOffice.org 1.1.3

### Limitations of the New Spell Checker

The new Kiswahili spell checker currently has only 4894 stems in the wordlist which were not enough for the spell checker to check all the existing Kiswahili words.

Both prefixes and suffixes were created based on noun classification and sampling some Kiswahili verb stems that followed the same trends to make other valid words as it was described in Affix Creation section. These affix rules might not be applicable to some Kiswahili verb stems that were not in the wordlist.

The spell checking system MySpell which was used to develop the Kiswahili spell checker is limited to appending the affix to the stem; therefore, newly discovered words should be manually added to the wordlist.

### CONCLUSIONS AND FUTURE WORK

Kiswahili is the fast growing language in Africa and the rest of the world, efforts have begun to localize available English software products into Kiswahili language. The most useful software products are office suites and web

browsers. Currently, there is Kiswahili office called Jambo Open Office and browser called Jambo Mozilla Firefox which were obtained after localizing free and open source software products called OpenOffice.org and Mozilla Firefox. To make these software products useful to Kiswahili users, a new spell checker has been developed to overcome the analyzed shortcomings of the existing Kiswahili spell checker for these products, and a further research will be carried out to create standard Kiswahili morphological rules and localizing more English software products into Kiswahili language. Currently, the Kilinux project is upgrading the Jambo OpenOffice.org 1.1.3 to version 2.0, and Jambo Mozilla Firefox 1.0.3 to 3.0. These Kiswahili software products will use the developed Kiswahili Spell checker.

## REFERENCES

- Christian, M, Lori, L, Rodolfo V, Ralf, B, Ariadn, FL, Alon, L, Jaime, C, Eliseo, C& Rosendo, H 2004, 'Data Collection and Analysis of Mapudungun Morphology for Spelling Correction', paper presented *In proceedings of the fourth international Conference on Language and Evaluation (LREC)*, Lisbon, Portugal. Viewed on 17 March 2008. <http://www.cs.cmu.edu/~alavie/papers/Mapudungun-LRE04.pdf>.
- 'Institute of Kiswahili Research (IKR)' 1930, viewed on 6 January 2007. <http://www.udsm.ac.tz/ikr/>,