

OpenOffice.org: Build Dictionary and Spellchecker in Your Language!

Kartik Mistry

Magnet Technologies Pvt. Ltd.

Mumbai



Little about me

- Gujarati Language Translator
- Native Language Lead! <http://gu.openoffice.org>
- Aspell-gu Author

The road ahead

- Basic requirements
- Dictionary/Wordlist: where to get it?
- Glance at Hunspell dictionary
- Building procedure
- When life is easy..
- Demo and conclusion

Requirements

- Linux/Windows/Mac machine
- Wordlist in your language, Not there? Wait!
- Courage to build OO.o (i.e. 2.3.x) on your machine. Or do you don't want that too? ;)

WordLists

- .. is simple word collection in alphabetical dictionary order according to your language
- Probably, you have it for your language! <http://aspell.net>
- Missing is: Kannada (ka), Any takers?

WordLists

- OpenOffice Dictionaries ready made at,
<http://wiki.services.openoffice.org/wiki/Dictionaries>
- Look for *Gora Mohanti* at IndLinux Project Day! You will get better idea!

Converting To Hunspell

- Hunspell supports Unicode very well, so its OO.o's choice
- Install it, <http://hunspell.sourceforge.net/>
- Lets look at sample file,
 - a. Wordlist file
 - b. Affix file
- Possible License Issue(s)

Before Building..

- Make sure that you have,
enough time/cpu power/passion
bandwidth
- OO.o build guidelines at, <http://tools.openoffice.org/>
- We need some simple tweaks!

Little tweaks!

- `./configure` with `--enable-hunspell`
- Dictionary files into `src/dictionaries/gu_IN/`
- `makefile.mk`
- `dictionary.lst`

Little tweaks!!

- `src/dictionaries/prj/build.lst`

```
di dictionaries\gu nmake - all di_gu di_diclst NULL
```

- Look in other provided dictionaries when in doubt
- Use sane editor!

Tired of compiling..

- .. Life is easy!
- DictOO.o wizard
- Demo
- Zipped dictionary

Little bit demo of my OO.o!

Resources

- Lingucomponent Project:
<http://lingucomponent.openoffice.org/dictionary.html>
- Mailing List: dev@lingucomponent.openoffice.org
- Hunspell Project: <http://hunspell.sourceforge.net/>

Questions



Thanks!



kartik_mistry@openoffice.org

<http://gu.openoffice.org>

Useless Stuffs: kartikmistry.org/blog