

Hacking with the Language

An Introduction to Natural Language
Computing

Santhosh Thottingal

Agenda

What is language computing and its importance

L10N and I18N

Applications for language

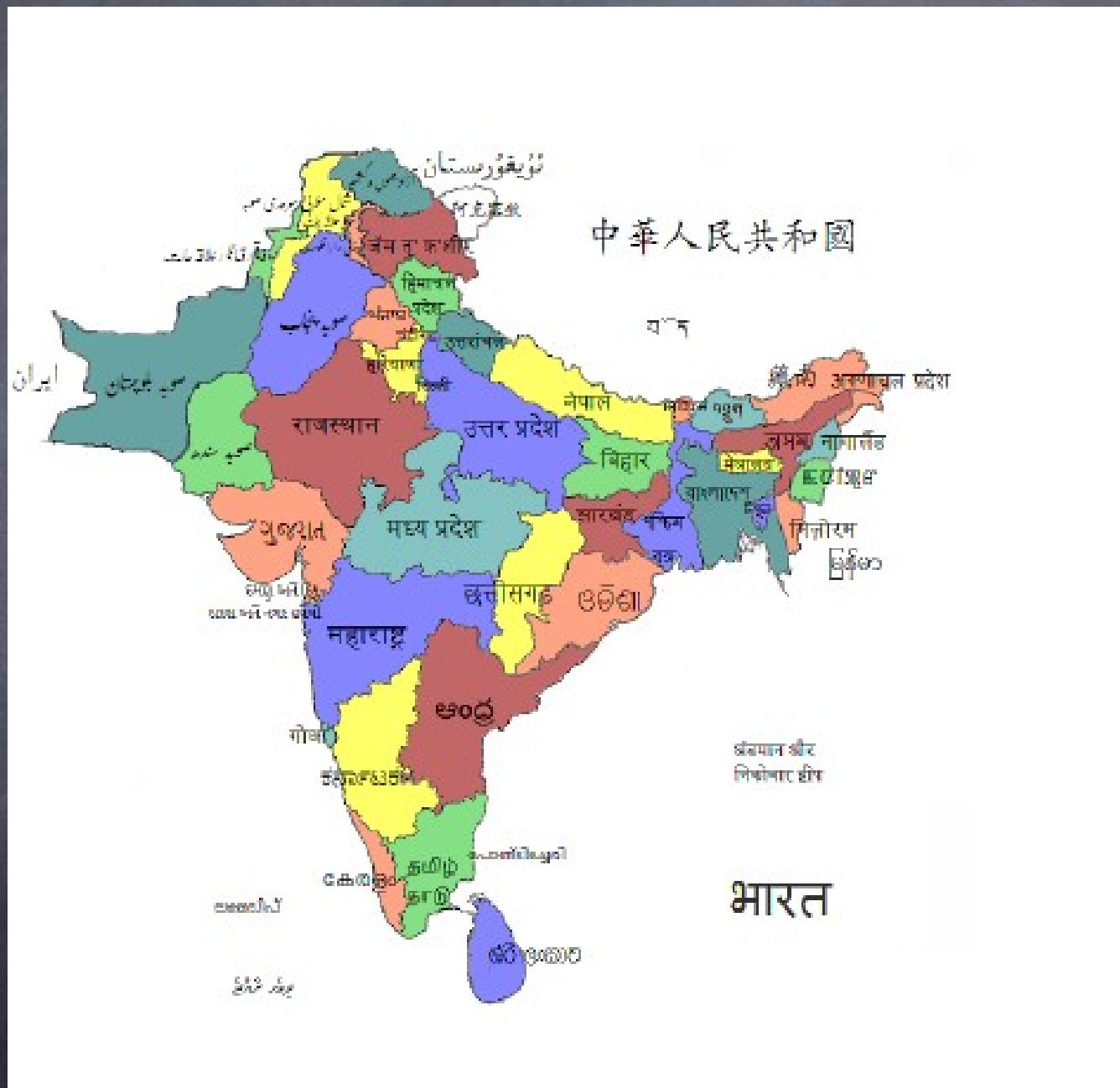
Challenges

Contributing

Why?

- English- De facto language?
- English covers 5-15% of word population
- Preserving our data in our own language
- Not only preserving , processing too!
- Technology for all- removing the language barriers
- Effective communication.

Languages



Digital Representation-Unicode

അയ്യപ്പൻ്റെ നെയ്യപ്പം ചുട്ടു

=\u0d05\u0d2f\u0d4d\u0d2f\u0d2a\u0d4d\u0d2a\u0d28\u0d4d\u0d31\u0d2e\u0d4d\u0d2e\u0d28\u0d46\u0d2f\u0d4d\u0d2f\u0d2a\u0d4d\u0d2a\u0d02\u0d1a\u0d41\u0d1f\u0d4d\u0d1f\u0d41

Rendering

ക + േ = കേ

ഭ + ു = ഭൂ

ഭ + ൠ + യാ = ഭൂയാ

ഗ് + ഭ് + യ് + രോ = ഭ്രോഗ്

क + ि = कि

Rendering Engines:

Pango,

Qt,

HarfBuzz,

Uniscribe(MS)

Fonts

" I can't go to a restaurant and order food because I keep looking at the fonts on the menu." - Donald Knuth

തപ്പോ തപ്പോ തപ്പാണി

- Anjali

തപ്പുകുടുകയിലെത്തുണ്ട് ?

- DYUTHI

നാഴിയുരി ചോറുണ്ട് !

- Meera

അമ്മാമൻ വന്നെ വിളമ്പാവൂ !

- Rachana

അമ്മാമി തന്നെ ഉണ്ണാവൂ !

- Suruma

Internationalization

C Example

```
AtkObject *obj;  
obj = gtk_widget_get_accessible (save_button);  
atk_object_set_name (obj, _("Save"));  
atk_object_set_description (obj, _("Saves all the settings and  
dismisses the dialog."));
```

But it is English...?!

Localization

ଅନ୍ତର୍ଜାତୀୟ

ସଂଗ୍ରହ

സംരക്ഷിക്കുക

சேமி

ସଞ୍ଚୟ

सहेजें

ਸੰਭਾਲੋ

దాచు

....

_(Save) =

Free software and Language Computing

Transparency

Developed and maintained by people from their
language

Language is no ones property

Applications

Input Methods

spellcheckers

Text to speech system

speech recognition

Grammer checkers

Optical character readers

Dictionaries

Automated Translation

Data mining

Natural Language Processing

What you can do?

1. Participate in Localization projects.

A stepping stone to foss development

familiarization of community based s/w development

Good understanding of FOSS applications and environments

Bringing the benefit of IT to commons

2. Develop Internationalized Applications- Not a feature, but requirement

3. Develop applications for the language – Don't expect that foreigners will do for us to consume!

How?

- Be a member of any language computing community. For eg: SMC for Malayalam
- Learn to type in your mother tongue!
- Participate in localization projects
- There are many projects waiting for developers- speech recognition, OCR, NLP related projects, calendars..
- Internationalize existing applications
- Google Summer of code 2009 is coming...!

Q&A

Santhosh Thottingal (<http://santhoshtr.livejournal.com>)

"എന്റെ കമ്പ്യൂട്ടറിന് എന്റെ ഭാഷ"

സ്വതന്ത്ര മലയാളം കമ്പ്യൂട്ടിങ്ങ് <http://smc.org.in>