



Creating Language Support Architecture On The Desktop For New Languages

- Rahul Bhalerao**
- Pravin Satpute**



The Plan..

- General Introduction
- The Detailed View
- Implementation on Desktop

Importance of Language Technology

- Languages as source of information and knowledge
- Practical Applications
- General misconceptions
- Statistics
 - India: 22 Officially recognized Languages
 - Computing possible for only 11 of them
 - English literacy and its need

What is Internationalization (i18n)?

Internationalization is the process of making software adaptable to the requirements of different native languages, local customs and character string encoding.



What is Localization (l10n)?

Localization is the process of adapting text and cultural content to specific target audience in specific location.

Language Technology beyond UI Translations..



The Architecture

l18n is not a feature, its an architecture!



I18n Components

Input

Output

Processing

Getting a new Language ready..

- Research on Language/Script
- Character Encoding
- Fonts
- Shaping Engines
- Input Methods
- OS Level Support



Character Encoding

- Encoding Standards
- Unicode
- Characters not in Unicode!



Fonts

- Role
- Creation
- Script Behaviors

Shaping Engines

- Role & interaction with fonts
- Support for various scripts
- The available ones

Input Methods

- The Mechanism
- Various options
- The Implementation
 - SCIM
 - M17n
 - Onscreen



Support on OS level

- Locale Definition
- Desktop Integration



Locale Definition

A locale is the definition of the subset of a user' s environment that depends on language and cultural conventions.

Categories of locale

- LC_CTYPE
- LC_COLLATE
- LC_MONETARY
- LC_NUMERIC
- LC_TIME
- LC_MESSAGES

- Some shortcut method to create and test Locale file
 - Python method
 - Adding locale file in archive

Desktop Integration

■ Gnome Desktop

- Gdm Language selection list
- Conditions
 - UTF-8
 - Iso-codes
 - Locale
 - Fonts & fontconfig
 - Translations



More Enhancements..

- Accessibility
- IO methods
- Machine Translations
- Character recognitions
- Language tools



Thanks..

Q?