



Linguistic Localization Framework for OOo

Jaganadh.G

Linguist

HDG – LTS

C-DAC (T)

jaganadhg@gmail.com



Overview

- Introduction
- Localization
- Linguistics and Localization
- Linguistic Localization
- Linguistic 110n framework for OOo
- Linguistic 110n in India



Introduction

- ❑ ICT Growth and Language Barrier
- ❑ Knowledge divide
- ❑ Digital Divide



Localization

- Types of Localization
 - Text Localization
 - Cultural Localization



Localization (l10n)

- Localization – l10n
 - Implementation of a specific language for an already internationalized software.
 - Adapting a program to a given culture
- Cultural Parameters
- Language rules
- Script – character set
- Date , time , currency
- Graphics & Icons



Resources for 110n

- Need locale info
- Collation sequence
- Fonts for proper rendering
- Input methods
- Language translations
 - GUI & Documentation



Linguistic localization

- Linguistic Localization is the process of localizing lingua components of selected software. Spell checker, Grammar checker Thesaurus etc are the lingua components.
- *Translation of interface and use of words*



Task evolved in Linguistic 110n

- Linguistic Standardization
- Testing and Evaluation of Localized applications
- Decision making
- Semiotics of 110n (Cultural 10n)
- Vetting of translated strings for a software
- Assisting the development of Lingua Components
- Evaluation of Translation



Linguistic Standardization

- Generating Standard Vocabulary
 - User friendly vocabulary
 - Selecting standard dialect
 - Following standard spelling
 - Omitting truncation errors



Decision Making

- ❑ Word to be translated and transliterated
- ❑ Standard transliteration
- ❑ Testing the translation



Semiotics of 110n

- Semiotics the study of signs and symbols.
- Selecting symbols and icons for a region
 - Culture neutral & self explanatory symbols for 110n



Lingua Component Development

- Development of Spellchecker.
- Generating Heuristic rule base for spellchecker,
- Incorporating features of selected languages in spellchecker.
- Generation of Lexicons
- Vetting of lexicon
 - Different parameter for lexicon generation & vetting
- Advanced NLP techniques for rule base development
- Context Sensitive spell checking



Lingua Components in OOo

- Spellchecker
- Grammar Checker
- Thesauri

- More to Add
 - Text to Speech
 - OCR
 - Integrated Input Method
 - ASR



Spell Checker

- ❑ Development of heuristic rule base for spellchecking
- ❑ Development of Spellchecker dictionary
- ❑ Language Specific Issues
- ❑ Context Sensitive Spell checking



Grammar Checker

- Data Generation techniques for Grammar Checker Development
- Corpus processing Techniques for Grammar Checker
- Parser based Grammar Checker (Link Parser)
- Rule based Grammar checking (Gravix etc....)



Thesauri

- Building Indic language Thesauri for Indic Language
- Selection of Lexicon preference
- Methods for finding lexicon preference



Adding More

- TTS System
 - Development of TTS system based on SSML

- OCR System Integration With OOo

- Translation System with OOo



Linguistic 110n in modern context

- Penetration of IT education up to grass root level
- Meeting needs of extreme level end-user
- Changing localized OOo as a National Standard for India



More Semiotics and Technology

- Selecting Region Specific Icons and Clip Arts
- Templates for layman Offices and Students (Templates for all)



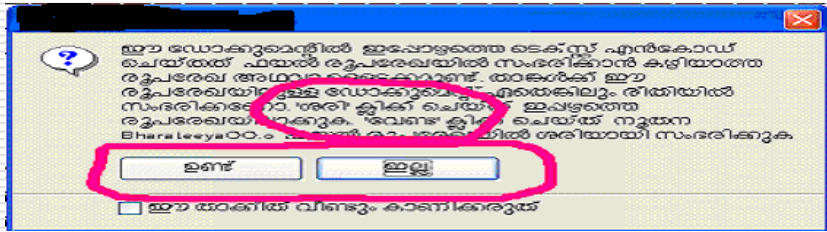
Community development for Linguistic 110n of OOo

- Building Community for Linguistic 110n
 - Community for languages
 - Inviting all to the community.



Some Examples







-
- ❑ **FURTHER READINGS**
 - ❑ **Anthony Pym, Localization and Linguistics, Intercultural Studies Group, Universitat Rovira i Virgili, Tarragona, Spain.**
 - ❑ **_____ , Localization: On its nature, virtues and dangers, Intercultural Studies Group, Universitat Rovira i Virgili, Tarragona, Spain.**
 - ❑ **_____ , Localization, Training, and the Threat of Fragmentation, Intercultural Studies Group, Universitat Rovira i Virgili, Tarragona, Spain.**
 - ❑ **_____ , What Localization Models Can Learn From Translation Theory Intercultural Studies Group, Universitat Rovira i Virgili, Tarragona, Spain.**
 - ❑ **Sasikumar (et.all), Guide to Free and Open Source Software Localization, CDAC, Mumbai.**
 - ❑ **< www.lisa.org > Localization Industry Standard Association Home Page**



-
- Question ????
 - Discussion



□ Thank You