Selected Text Machine Translator for English to Sinhala
Overview

• Introduction
• What is Machine Translation system?
• Why Selected text Translation?
• Design
• Implementation
• Evaluation
• Conclusions and Further works
Introduction

- Electronic documents are increasingly becoming popular.
- Many of such documents are available on the Internet in English Language.
- Entire world community in not conversant in reading the documents published in English. (Language barrier)
- Machine Translation systems are used to overcome this **language barrier**
Machine Translation

• Machine translation is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another (Wikipedia)

• Approaches
  - Human-assisted
  - Rule-based
  - Statistical
  - Example-based
  - Knowledge-based
  - Hybrid
  - Agent-based
Existing Machine Translation systems

Input sentence.
Machine translation is a sub-field of computational linguistics

Layered output

Open in New window
Why selected text Translation?

• Enables reader to get translation of selected texts without disturbing the process of reading.
Design

English to Sinhala Selected text translator

- BEES Client
- BEES Server
- BEES Translator
BEES-Translator

• English Language System
  – Morphological analyzer
  – English Parser
• English to Sinhala Translation system
• Sinhala Language System
  – Morphological generator
  – Syntax Generator
English Morphological Analyzer

- Reads a given English text word by word and identifies morphological information on each word.
- Customized an existing English morphological analyzer.
- Implemented using SWI-Prolog.

**Input English word:** Boy

**Category** | **Data**
---|---
Type | Noun
ID | E1000006
Person | 3rd
Number | Singular
Sex | Male
Case | Dative
Word | Boy

```
eng_noun([e1000006], td, sg, ma, sb, 'boy').
```
English Parser

- Works as a sentence analyser
- Analyses the English sentence or a phrase
- Return the grammatical information
  - Subject
  - Verb
  - Complement
- Customized an existing English parser
- Implemented using SWI-PROLOG
Base-word Translator

- Translate English base word into Sinhala base word with the help of bilingual dictionary.
- Use word to word translation at this level.
- Implemented using SWI-PROLOG
Sinhala Morphological Generator

- Uses concepts of Varanegeema
- Reads the words from translator word by word
- Generates appropriate Sinhala terms according to grammatical information
- Implemented using SWI-Prolog.
Sinhala Syntax Generator

- Works as sentence composer
- Receives tokenized words from the Sinhala morphological analyzer and composes grammatically correct Sinhala sentence / phrase.
- Implemented using SWI-PROLOG
Lexical Dictionaries

- Designed as prolog databases
- **English Dictionary**
  - Regular Nouns, Regular Verbs, Irregular nouns and irregular verbs are stored separately
- **Sinhala Dictionary**
  - Regular Nouns, Regular Verbs, Irregular nouns and irregular verbs are stored separately
  - Rules are also included
    - Nama Gana, Kriya Gana, Vibakthis forms
- **English-Sinhala Bilingual Dictionary**
  - Used to identify appropriate Sinhala base word for the given English word
  - Shows relations between English and Sinhala words
Evaluation

• System has been tested through the human support
• Same source (sample text file) is given for the five users
• Randomly selected text has been used to translate
• Quality of the translation is tested through the knowledgeable people
• Ranks it as good, fair, accept, week or error
• Evaluation is done for the 50 times
• Number of word in the randomly selected text and result of the translation are recorded
System gives more accurate results for the multiple word (2 or 3) selection than a single word selection.
Conclusions

- Translate selected text approximately 80% accuracy
- System gives more accuracy for the limited number of selected words
- **Issues** - Incorrect text selection
- English parser and the Sinhala sentence generator in the translation system are needed to improve to handle incomplete sentences
- Research has now been conducted to use the Multi-agent technology to improve the performances
Thank you!