Computational Linguistics (CL3.101)

Summer 2021, IIIT Hyderabad

01 June, Tuesday (Lecture 4)

Taught by Prof. Radhika Mamidi

Applications

Spell checkers, grammar checkers and word predictors are a common application of these technologies.

Machine Translation

It is a subfield of computational linguistics that uses computer programs to automatically translate from one language to another.

A common approach is to translate from the source language to an abstract formal representation, and then to the target language. This allows us to avoid creating a new translation for each pair of languages.

Methods

1. Rule-based
2. Corpus-based This method uses statistics to find equivalents for a word, phrase or sentence from the parallel corpus. The program learns from aligned parallel corpora.

Types of MT Systems

1. Dictionary-based
   This follows the word-for-word translation method, using a dictionary to translate each word.

2. Statistical
   This system follows the corpus-based method.

3. Example-based
   This essentially follows a translation-by-analogy approach.

4. Neural networks This model predicts the likelihood of a sequence of words, modelling entire sentences in a single model.

Limitations

- It can't translate cultural components of the text.

- It often generates overly literal output.
- It cannot recognise idioms or slang if they are not in its memory.
- It lacks the creativity of human work.