

## Use of dynamic text prediction for Tamil text editing

Sarubi T., Lorensuhewa S.A.S. and Kalyani M.A.L.

*Department of Computer Science, University of Ruhuna, Matara, Sri Lanka.*

In text processing, typing whole documents by ourselves leads to lots of spelling mistakes and also it is time-consuming. When it comes to morphologically rich languages like Tamil language, it's even more difficult, due to the absence of a clear picture of the Tamil keyboard layout. The aim of this research study is to develop a user-friendly tool to perform next word prediction and spell checking. In our approach, while user types, we detect the user's typing domain using a classifier and then predict the next word according to the predicted domain. Next word prediction is done using domain-specific language models by giving priority to trigram and then bigram. Language models can continuously learn from user's typing. Recency-based model is used to reduce the search space. Also, detect misspelt words and propose dictionary lookup with distance measure and improve the dictionary suggestion list using n-grams lookups. According to our experiments, Tamil language results in lowest word prediction percentage (WPP) accuracy among Sinhala and English languages. We further analyzed results by varying the total number of words in all three languages and counted the number of unique words. It can be seen from the results that Tamil language has the highest unique words compared with the other two. Tamil language has a large vocabulary than the other two languages and we believe that the lowest prediction level was obtained due to this diversity. Dynamic prediction helps the users, because within a document, we may need different domain n-gram models to predict words. Dictionary lookup with forward and backward bigrams show highest improved accuracy, of 54% while dictionary lookup achieved 36% accuracy.

**Keywords:** Tamil language, dynamic word prediction, language models, spell checker

*\*Corresponding author: stsarut@gmail.com*