

## Lexicon splitting in lexical disambiguation for Malay morphological analysis and stemming

### Abstract

Lexical ambiguity is one of the problems faced by morphological analyser and stemmer. It is caused by ambiguous word form like homonym, which could direct the tools to produce incorrect output. Thus a method that can resolve ambiguity may improve the performance of such tools. Malay word affixation differentiates between monosyllable and multisyllable word. A disambiguation method is proposed for tools that use lexicon for analysis and stemming, by splitting the lexicon into monosyllable and multisyllable words. We found that this feature could help to resolve ambiguity involving monosyllable words, improve language's exception handling and improve storage lookup. This would be useful for Malay morphological analysis and stemming as this method does not require document-level context analysis of the analysed word.

**Keyword:** Natural language processing; Morphological analysis; Stemming; Malay morphology; Lexical disambiguation