

Corpus Analysis: A Case Study on Kadazandusun Newspaper Archive

ABSTRACT

This paper presents the analysis of text data acquired from News Sabah Times, which is the only Sabah's newspaper that has a section for Kadazandusun news. Currently, there is no text translation tool available to translate a sentence or large text specifically from Kadazandusun to other languages such as Bahasa Melayu. Thus, the first step is to develop such a system is to analyze the available corpus from the newspaper archive. The objective is to perform text analysis and then providing the possible ways of utilizing the knowledge. In this work, the purpose is to report the methodology and utilization of the fundamental corpus analysis related to text mining and not covering on the linguistics aspects and grammatical context. In addition, this paper also reports on the findings from the newspaper corpus analysis.