

Developing an open source Kipsigis spell checker and language tool

Ronoh, Wycliffe

Abstract:

This report discusses the morphology of the various parts of speech of the Kipsigis language, the development of corpus preparation and analysis tool using Jython and describes the development of an open source Kipsigis spell checker using the Hunspell language tools. Kipsigis is a resource scarce language whose print and digital usage is low; developing a spelling recognition system requires a lot of manual effort on corpus preparation and analysis. We describe the development of a tool that helps to automate and thus speed up this procedure. Hunspell requires two files to define the language that it is spell checking, the first file is an affix file that defines the meaning of flags, the second file is a dictionary file which contains words alongside flags. The spell checker tested on four data sets ranging from 460 to 540 words achieves an average accuracy rate of 96%, an average precision rate of 100%, an average recall rate of 95% and an average coverage rate of 94%. The spell checker developed seeks to be adopted by leading open source systems such as Open Office, Mozilla and Google chrome.