

Masood Ghayoomi & Ehsan Daroodi

Using POS Tags in Word Prediction: A Statistical Language Modeling for the Persian Language

This paper discusses word prediction systems developed for the Persian language. The systems use statistical syntax along with word statistics to increase the accuracy of predictions. Main syntactic categories are used in one system to make predictions; and another system has used the main syntactic categories along with their morphological, syntactic and semantic subcategories in predictions. After a brief introduction and review of related literature and systems, the authors compare the systems with the other system which solely uses word statistics to make predictions. Three standard performance metrics are used in this respect, including keystroke saving, the most important one. The results show the designed system using POS tags, of both main syntactic categories and their subcategories, achieved over 42% saving in keystrokes, the highest percentage; also the performance of the system is significantly better than the other systems which use merely word statistics, and word and POS tags of the main syntactic categories.