Towards Arabic Spell-Checker Based on N-Grams Scores

Abstract

The main purpose of this paper is to develop a simple and flexible spell-checker for Arabic language. The proposed spell-checker is based on N-Grams scores. For this purpose, eleven matrices are built to present the combination between the Arabic letters word. Each matrix concerns in the connection between a 2-grams letters. Each cell in the generated matrix is assigned an integer value 2, 1 or 0. The cell is assigned the value 2 in the corresponding matrix; if the word is ended by these two letter and assigned 1 if there is a connection and the word is not over yet, and is assigned 0 otherwise. On the other side searching process for any word that is by extracting each pair of letters in the word then it examines the value for each pair when the corresponding value is zero then the spell checker will consider the test word as wrong; otherwise it will check if it is assign with 1 that indicates that there is a connection it will be continue until reach to the value of 2 to determine that the word is correct. The overall accuracy for the proposed spell-checker is reached to 98.99%.

References

- Feldman A. Computational linguistics: Models, resources, applications. Computational
Towards Arabic Spell-Checker Based on N-Grams Scores


Index Terms

Computer Science Artificial Intelligence

Keywords

Natural Languages Processing Arabic Language Processing Spell-Checker N-Gram