Presentation Topic:

Arabic Text Correction Using Dynamic Categorized Dictionaries: *A Statistical Approach*

Presenters:

Prof. Adnan Yahya & Mr. Ali Salhi

Abstract:

In our presentation we will describe a technique for spelling and correcting Arabic text that provides different variables that can be controlled to give customized results based on the properties of the processed text. The proposed technique depends on dynamic dictionaries controlled and customized based on the input text categorization. In the research reported here we employ a statistical/corpus-based approach with data obtained from the Arabic Wikipedia and local Palestinian newspapers. Based on corpus statistics we constructed databases of words and their frequencies as single, double and triple expressions plus dynamic dictionaries based on categorized text and used that as the infrastructure for our spelling and text correction technique. We briefly report on the results of preliminary testing and analysis. While the results reported here are promising, they must be viewed as work in progress, still in need of more testing, refining, integration and deployment in real life settings.

Please Note:

The presentation will be based on early work we did and presented in the 4th International Conference on Arabic Language Processing, May 2–3, 2012, Rabat, Morocco.

Our paper with title *(Arabic Text Correction Using Dynamic Categorized Dictionaries: A Statistical Approach*) was presented in the mentioned conference and was selected for publication in the International journal of Computational and General Linguistics: LINGUISTICA COMMUNICATIO, Vol. 5, 2013.