

Integration of an Arabic Lemmatizer as Part of an Information Retrieval System



Mohamed Amine Cheragui¹, Zoulikha Benblal¹ and Fatima Belouafi¹

✉m_cheragui@univ-adrar.edu.dz

¹LDDI Laboratory, Mathematics and Computer Science Department, University Ahmed Draia of Adrar, National Road N 6, Adrar 01000, Algeria

Received: 10 October 2021

Accepted: 20 November 2021

Published online: 31 December 2021

Abstract

Natural Language Processing (NLP) is a field of Artificial Intelligence that encompasses several topics: Machine translation, Text summarization, spell checkers, information retrieval, etc. The objective of information retrieval (IR) is to provide a user with easy access to the information of interest, this information being located in a mass of textual documents. To achieve this objective, it must represent, store and organize information, then provide the user with the elements corresponding to the information need expressed by his request. Most of information retrieval systems (IRS) use simple terms to index and retrieve documents based on models such as: Boolean model, vector model, probabilistic model, language model, etc. The aim of our work is to develop a hybrid Arabic system based on four (04) models where the originality of the work lies in incorporating a stemmer in the search process to improve the results of our system.

Keywords Arabic language, Natural Language Processing, Information Retrieval, Stemming.