Abstract: This paper provides an improvement to Arabic Information Retrieval Systems. The proposed system relies on the stem-based query expansion method, which adds different morphological variations to each index term used in the query. This method is applied on Arabic corpus. Roots of the query terms are derived, then for each derived root from the query words, all words in the corpus descendant from the same root are collected and classified in a distinct class. Afterward, each class is reformulated by co-occurrence analysis (for each pair of terms) to guarantee a good relationship between terms. In addition, we used the whole word indexing technique to index the corpus. Each term added to the query must be correct and correlated to the original query terms. We show how this technique improves recall (by 20%), and how it affects precision.

Keywords: Arabic query expansion; Query expansion; Arabic stemming.