

# Jordan University of Science and Technology

## Automatic Extraction of Arabic Multi-Word Terms

**Authors:** Khalid Al Khatib and Amer Al-Badarneh

**Abstract:** Whereas a wide range of methods has been conducted to English multi-word terms (MWTs) extraction, relatively few studied have been applied to Arabic MWTs extraction. In this paper, we present an efficient approach for automatic extraction of Arabic MWTs. The approach relies on two main filtering steps: the linguistic filter, where simple part of speech (POS) tagger is used to extract candidate MWTs matching given syntactic patterns, and the statistical filter, where two statistical methods (log-likelihood ratio and C-value) are used to rank candidate MWTs. Many types of variations (e.g. inflectional variants) are taken into consideration to improve the quality of extracted MWTs. We obtained promising results in both coverage and precision of MWTs extraction in our experiments based on environment domain corpus.