

# Jordan University of Science and Technology

## A comparison study of some Arabic root finding algorithms

**Authors:** Al-Shawakfa Emad, Al-Badarneh Amer, Shatnawi Safwan, Al-Rabab'ah Khaleel, & Bani-Ismael Basel

**Abstract:** Arabic has a complex structure, which makes it difficult to apply natural language processing (NLP). Much research on Arabic NLP (ANLP) does exist; however, it is not as mature as that of other languages. Finding Arabic roots is an important step toward conducting effective research on most of ANLP applications. The authors have studied and compared six root-finding algorithms with success rates of over 90%. All algorithms of this study did not use the same testing corpus and/or benchmarking measures. They unified the testing process by implementing their own algorithm descriptions and building a corpus out of 3823 trilateral roots, applying 73 trilateral patterns, and with 18 affixes, producing around 27.6 million words. They tested the algorithms with the generated corpus and have obtained interesting results; they offer to share the corpus freely for benchmarking and ANLP research.