

# Jordan University of Science and Technology

## Automatic Keyphrase Extractor from Arabic Documents

**Authors:** Hassan M. Najadat, Ismail I. Hmeidi, Mohammed N. Al-Kabi and Maysa Mahmoud Bany Issa

**Abstract:** The keyphrase is a sentence or a part of a sentence that contains a sequence of words that expresses the meaning and the purpose of any given paragraph. Keyphrase extraction is the task of identifying the possible keyphrases from a given document. Many applications including text summarization, indexing, and characterization use keyphrase extraction. Also, it is an essential task to improve the performance of any information retrieval system. The internet contains a massive amount of documents that may have been manually assigned keyphrases or not. The Arabic language is an important language in the world. Nowadays the number of online Arabic documents is growing rapidly; and most of them have no manually assigned keyphrases, so the user will scan the whole retrieved web documents. To avoid scanning the entire retrieved document, we need keyphrases assigned to each web document manually or automatically. This paper addresses the problem of identifying keyphrases in Arabic documents automatically. In this work, we provide a novel algorithm that identified keyphrases from Arabic text. The new algorithm, Automatic Keyphrases Extraction from Arabic (AKEA), extracts keyphrases from Arabic documents automatically. In order to test the algorithm, we collected a dataset containing 100 documents from Arabic wiki; also, we downloaded another 56 agricultural documents from Food and Agricultural Organization of the United Nations (F.A.O.). The evaluation results show that the system achieves 83% precision value in identifying 2-word and 3-word keyphrases from agricultural domains.