

# Jordan University of Science and Technology

## A Survey of Text Categorization Techniques Using Arabic Text

**Authors:** Ismail Hmeidi, Amer Al-Badarneh, Haya Rababah, Noor Abuelrub, and Dua'a Alawad

**Abstract:** Text classification (TC) is one of important techniques used to categorize documents into a predefined categories based on their content. It is used in many applications such as classification of web pages, email messages, and news stories. Recently, there are some research works on applying TC techniques using Arabic text. The main goal of this paper is to review and compare some of the different approaches of existing TC algorithms using Arabic text. The authors are not aware of such study that has been published in literature yet. To conduct this study, we have searched for published algorithms that are recent and popular and have higher rates of success in classifying Arabic text. The authors have studied and compared 15 recent TC techniques applied to Arabic data. All techniques of this study did not use the same testing corpus and/or benchmarking measures.