

Jordan University of Science and Technology

A New Efficient Kea Based Approach for E-Mail Spam Detection

Authors: Qusai Abuein, Hassan Najadat, Sana Wedian and Deya Alzoubi

Abstract: With the increased advancement in technology and the proliferation of internet applications, electronic mails become an increasingly essential means of communication, for both individuals and organizations. In the recent years, however, spam's become the most unsolicited forms of messages that invade the most important services of internet: Electronic mail and search engines. The last decade witnessed a great reaction towards these annoying spam's by applying a variety of filtering techniques to combat spammers, based on the assumption that in any spam mail, there are specific words/patterns that provide indications of spam's, that is, there are predefined barriers that distinguish a spam message from a legitimate one. The most used techniques for spam detection are Bayesian Classifiers that have been argued to be efficient for filtering E-mail spam. However, as spam detection techniques evolve, spammers evolve too so as to prove their excellence to adapt to the predefined barriers. Therefore, content- based recognition techniques must be applied on e-mails to detect spams based on the semantic of their contents. In this paper, we propose a new approach for e-mail spam detection based on kmeans algorithm and the Keyphrase Extraction Algorithm (KEA). Our approach achieves high classification accuracy in the sense that it takes into consideration the semantic nature of the textual contents.