Abstract

The virtual world provides criminals with an anonymous environment to conduct malicious activities such as mal-ware, sending ransom messages, spamming, theft intellectual property and sending ransom e-mails. All these activities are text in somehow. Therefore, there is a need for a tool in order to identify the author or creator of this illegal activity by analyzing the text. Text-based Authorship Identification techniques are used to identify the most possible author from a group of potential suspects of text. This paper is meant to explore the text-based authorship identification researches within the period 2007-2017. The researches were classified based on the application into email authorship, source code authorship, online text authorship, gender identification and online messages authorship. Also, the paper reviews and reports the datasets which used in the experiments of text-based authorship identification techniques. Finally, it reported the techniques which were used in authorship identification.