High Capacity Steganography Tool for Arabic Text Using ‘Kashida’

Adnan Abdul-Aziz Gutub a,∗ and Ahmed Ali Al-Nazer b

a College of Computer, Umm Al-Qura University, Makkah, Saudi Arabia.
b Saudi Aramco, Dhahran, Saudi Arabia.

A R T I C L E I N F O

Article history:
Received: 29 November 2009
Revised: 9 June 2010
Accepted: 16 June 2010
Published Online: 13 July 2010

Keywords:
Arabic E-Text, Text Steganography, Text Watermarking, Text Hiding, Kashida, Feature Coding

A B S T R A C T

Steganography is the ability to hide secret information in a cover-media such as sound, pictures and text. A new approach is proposed to hide a secret into Arabic text cover media using “Kashida”, an Arabic extension character. The proposed approach is an attempt to maximize the use of “Kashida” to hide more information in Arabic text cover-media. To approach this, some algorithms have been designed and implemented in a system, called MSCUKAT (Maximizing Steganography Capacity Using “Kashida” in Arabic Text). The improvements of this attempt include increasing the capacity of cover media to hide more secret information, reducing the file size increase after hiding the secret and enhancing the security of the encoded cover media. This proposed work has been tested outperforming previous work showing promising results.

© 2010 ISC. All rights reserved.

1 Introduction

Steganography is defined as in [1] “the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, even realizes there is a hidden message”. Steganography works as we hide information in un-used and redundant bits in any cover media such as pictures, sound and text.

Hiding secret information in text is more challenging. First, text documents have relatively little redundant information. Second, the structure of text documents is almost identical to their look and hence any change may be visible. Nevertheless, using text is preferred over other media because it needs less memory to save, is easier to transfer over the network and more efficient and cost-saving in printing [2, 3].

Text steganography as it is hiding a secret inside text has dependencies on the language used as cover media. Different human being languages have different characteristics and properties. In Arabic language, there are 28 different characters. Arabic characters are joined when writing words contain more than one character. Depending on the joined characters, an extension character “Kashida” may be embedded between two Arabic characters.

There are two uses of the extension character “Kashida” in Arabic text. One is to decorate the Arabic text format so that it looks better and more convenient. This use is important especially in the titles of the documents. The second use is to justify the Arabic writings within lines, similar to English where spaces are used for justifying the text in lines. The advantages of using “Kashida” in Arabic text to either format it or justify the lines will not affect the text contents and meaning [4, 5].

In this paper, an improved approach is proposed to maximize the use of the Arabic extension character, Kashida, between joined characters in Arabic text cover media. The idea of this approach is to embed “Kashida” wherever possible after any Arabic letter regardless of it being dotted or not dotted; as Arabic