International Journal of Advanced Pervasive and Ubiquitous Computing
Volume 11 • Issue 1 • January-March 2019

A Novel Coding and Discrimination (CODIS) Algorithm to Extract Features from Arabic Texts to Discriminate Arabic Poems

Nada Ahmed J., MOP, Baghdad, Iraq

Abdul Monem S. Rahma, Computer Science Department, University of Technology, Baghdad, Iraq

Maha A. Hmmood Alrawi, Department of Production Engineering and metallurgy, University of Technology, Baghdad, Iraq

ABSTRACT

This article proposes a new algorithm to discriminate Arabic poems by inserting Arabic poems texts and coding Arabic letters, extracting letters features depending on letter shapes to construct a multidimensional contingency table, and analyses the frequencies of letters in the inserted texts statistically. The proposed coding and discrimination (CODIS) algorithm could be applied for different Arabic texts in any media. A sample of five poems for six poets was examined to implement a CODIS algorithm. A Chi-Square statistic is used to determine the relation between the features and discriminate poems.

KEYWORDS

Arabic Text, Coding, Letters Shapes, Poems

1. INTRODUCTION

In the modern information society, computers are used in relation to a very high proportion of communication. The Arabic language faces many challenges, due to the lack of researchers to support Arabic language, and the narrow field of Internet search using Arabic language as compared with English language and others. On the other hand, there has been no standard Arabic reference to solve Arabic letters problems (Saleh, 1994).

The aim of this work is to propose an algorithm to extract some features of the Arabic letters to facilitate dealing with any Arabic text in different media such as poems, written stories, manuscripts, books, and others as a general, and to find an efficient procedure to discriminate Arabic texts as a particular. The proposed coding and discrimination (CODIS) algorithm is applied to discriminate Arabic poems in different literature ages and could be a useful instrument to distinguish the Anonymous Poems. The coded letters could be used for different purposes too like programming languages and secure decoding systems to support the integration of Arabic language into other computer systems such as automated speech and automatic recognition of Arabic speech, search engines and others.

DOI: 10.4018/IJAPUC.2019010101

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

2. ARABIC POEMS AND THEIR STRUCTURAL CHARACTERISTICS

Arabic poems have been developed from classical poem to modernist poems in the second half of the 19th century. The poetic development and the new thematic features of poems have been documented and analyzed in the 20th century (Jansson, 2010).

Art of poem is one of the most popular literature arts, and the most widespread because oldest of humanity uses it. Poem is the first expressive literature picture that appeared in human life since Early times, this oldest of poem returned to necessary of biology life (Ismail, 2013).

Poem is a way that human can express about themselves like angry, joy, and sadness. Humans can express throw poem about their mind and reactions and ideas, therefore to poem flexible neutrality bailed according to desire of human itself (Al-Nakhal, 2013).

Weighted words in poem lines describe the quality of meanings and choosing vocabulary and Expression accuracy, and affect the emotion of peoples (Jansson, 2010). Many statistical features could be extracted from poems texts especially those are with weighted rhymes. The most efficient parametric variable is the *frequency* of each letter in the texts inserted to be coded by (CODIS) algorithm. These letters are categorized depending on different criteria which will be explained in the next section.

3. ARABIC LETTERS CLASSIFICATION AND GROUPS

Arabic language is one of the oldest Semitic languages; it is the mother language of vast number of individuals around the world. The Holy Quran is in Arabic language (Khallati, 2016). It is the gateway to the most culturally rich areas of the world, as a result of the important cultural changes that the Arab region has submitted. Historians need to learn the arts of this language so that they can communicate with the Arabs, and to carry out the enormous Arabic manuscripts and valuable books (Khallati, 2016).

Arabic language can be divided into two versions: classical Arabic and Modern Arabic. Classical Arabic is the official language and it is the language of the Quran. It is also used in newspapers, books, and academic researches. As classical Arabic is most used in written context, Modern Arabic is used in the daily spoken contexts (Eidus, 2007; Mrayati, 2004).

Arabic Alphabets consist of 28 letters. There is no doubt that the Arabic alphabet (أ ب ج د هـ و ز) have already been mentioned based on the alphabetical order (أ ب ت). However, there are many differences with regard to symbols or non-Arabic characters, and the additional letters have spaces between them such as ء ى ة)). All are not assembled on their number or on their positions in text.

Letters coding is one of the most important fields of the computer programs. Many studies dealt with this field in different systems to develop efficient secure systems (Salomon, 2007). Arabic letters was classified to groups; these groups are as bellow which categorized depending on different criteria:

A. Classification of Letters Depending on Letter Location in the Word.

Arabic letters forms have four groups classified according to their position in the word as first, middle, end, or separated letters (as letter's form):

- The first letters group is the letters in the forms: (بـ، تـ، ثـ، جـ، حـ، خـ، سـ، شـ، صـ، ضـ، طـ، ظـ، عـ، غـ، فـ، قـ، كـ، لـ، مـ، نـ، هـ، يـ، ئـ).
- The middle letters group is the letters in the forms: (بـ، تـ، ثـ، جـ، حـ، خـ، سـ، شـ، صـ، ضـ، طـ، ظـ، عـ، غـ، فـ، قـ، كـ، لـ، مـ، نـ، هـ، يـ، ئـ).
- The final letters group is the letters in the forms: (ـا، ـب، ـت، ـث، ـج، ـح، ـخ، ـد، ـذ، ـر، ـز، ـس، ـش، ـص، ـض، ـط، ـظ، ـع، ـغ، ـف، ـق، ـك، ـل، ـم، ـن، ـه، ـو، ـي، ـة، ـؤ، ـئ، ـأ، ـإ، ـى، ـآ).
- The separated letters group is the letters in the forms: (ء، ا، ب، ت، ث، ج، ح، خ، د، ذ، ر، ز، س، ش، ص، ض، ط، ظ، ع، غ، ف، ق، ك، ل، م، ن، ه، و، ي، ة، ؤ، ئ، أ، إ، ى، آ).

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-novel-coding-and-discrimination-codis-algorithm-to-extract-features-from-arabic-texts-to-discriminate-arabic-poems/224936

Related Content

Frequency Offset Estimation Algorithm of OFDMA Wireless Communication System Based on Compressed Sensing
Ma Qinggongand Yang Bo (2015). *International Journal of Advanced Pervasive and Ubiquitous Computing (pp. 55-62).*
www.irma-international.org/article/frequency-offset-estimation-algorithm-of-ofdma-wireless-communication-system-based-on-compressed-sensing/165179

A Software Engineering Perspective on Ubiquitous Computing Systems
Emerson Loureiro, Loreno Oliveira, Frederico Bublitz, Angelo Perkusichand Hyggo Almeida (2008). *Ubiquitous Computing: Design, Implementation and Usability (pp. 222-237).*
www.irma-international.org/chapter/software-engineering-perspective-ubiquitous-computing/30528

A Context-Driven Commit Protocol for Enhancing Transactional Services Performance in Pervasive Environments
Widad Ettazi, Hatim Hafiddiand Mahmoud Nassar (2018). *International Journal of Advanced Pervasive and Ubiquitous Computing (pp. 14-28).*
www.irma-international.org/article/a-context-driven-commit-protocol-for-enhancing-transactional-services-performance-in-pervasive-environments/211940

Data Association Based Tracking Traffic Objects
Tao Gao (2013). *International Journal of Advanced Pervasive and Ubiquitous Computing (pp. 31-46).*
www.irma-international.org/article/data-association-based-tracking-traffic-objects/93583

Analyzing User Behavior in Digital Games
Anders Drachenand Alessandro Canossa (2012). *Media in the Ubiquitous Era: Ambient, Social and Gaming Media (pp. 1-28).*
www.irma-international.org/chapter/analyzing-user-behavior-digital-games/58578