



7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

Flowchart Plagiarism Detection System: An Image Processing Approach

Jithin S Kuruvila, Midhun Lal V L, Rejin Roy, Tomin Baby, Sangeetha Jamal,
Sherly K K*

Rajagiri School of Engineering and Technology, Kakkanad, Kochi 682039, India

Abstract

Plagiarism is any identical or lightly-altered use of one's own or someone else's work. Text plagiarism detection systems are widely available. Even though image plagiarism detection systems exist, flowchart based plagiarism detection systems are rarely implemented. In the proposed flow chart plagiarism detection system, flowcharts are compared by comparing both the shape, orientation as well as text. This approach creates graph from the flowchart, hence is capable to detect the plagiarism with same shaped objects even though the orientation of the graph is changed. Accuracy of the method is tested with flowcharts of different shape and orientation.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications.

Keywords: Canny edge detection; Directed graph; Flow chart plagiarism detection; Flowchart preprocessing; Gaussian filter; Shape detection;

1. Introduction

Plagiarism is any alike or slightly modified use of one's own or someone else's work. Plagiarisms are mainly categorized as text based and image based plagiarisms. In text based plagiarism, the text written by an author is copied by someone else without citing the original author. In text plagiarism two cases can occur; the text written by a person will be copied as such by someone else or the text will be modified such that the words change but the

* Corresponding author. Tel.: +91-9495600124.

E-mail address: sherly.shilu@gmail.com

semantic meaning of the sentence remains the same. Plagiarism detection of first case is easy to detect while the later requires complex analysis and algorithms. Text plagiarism detection systems are widely available. In image based plagiarism the images that are drawn or created by a person will be taken by someone else and presented as their own work. In this type of plagiarism the image will be either taken as a whole or only a part of it will be taken. Even though image plagiarism detection systems exists, flowchart based plagiarism detection systems are rarely implemented.

A flowchart is a diagrammatic representation of an algorithm. It consists of various graphical shapes such as circle, rectangle, diamond, parallelogram etc. which are connected using connector lines. Each shape corresponds to a particular operation and have some text contained within it. The objective of this approach is to detect plagiarism in flowcharts. As the amount of information transferred by flowcharts is increasing every day, it is important to authenticate the flowchart.

2. Related Works

Currently, there are only a few researches working on the issue of image plagiarism, particularly diagrams and flowcharts. However, despite not many works are being presented in flowchart plagiarism system, there are some works related to this issue, such as methods to characterize flowchart types based on the image features. Sensory Arrish et. al. [1] present a method for detecting flow chart figure plagiarism based on shape-based image processing and multimedia retrieval. The method retrieve flowcharts with ranked similarity according to different matching sets. Vipul Bajaj et. al. [2] used a perceptual hash functions along with the rotation check to detect image plagiarism. Hermann Maurer et. al. [3] discusses the serious issues of plagiarism and the results of certain plagiarism detection software. L. McKeever [4] gives an overview of the currently available different methods of detecting web - based plagiarism. Miyao and Maruyama [5] used loop structure as source to detect directional strokes into learning system. The method managed to recognize 97.6 % of the given data sets, however it is still not fully practical as the system requires the user to draw a loop structure by himself.

A flowchart conveys information about process of works to the reader. Consequently, the text contained in the charts may be used as features of description. Vasudevan et al.[6] presented a method to extract information inside flowcharts by contour processing and neural network-based optical character recognition . The method analyzes flow lines and contour of the system as feature descriptors. The system managed to provide recall rate of 98%. Awal et al.[7] took a different approach to characterize flowcharts semantically using grammatical approach .

Aside from information contained in a flowchart, the shapes used by one node of flowchart is of equal importance, as it can distinguish between different processes in the work. It can be stated that flowchart is a subset of images, thus it can be characterized by kinds of image features. Furthermore, to develop a flowchart plagiarism system, the system should be able to recognize these features which will answer the required query by human. Zhang and Lu [8] reviewed techniques to represent and describe figures based on their shape features. The method depends on the feature inside each document like color, shape, and texture. In addition to color and textures, there are number of common representation of features vector, strings, and graphs, fuzzy and probabilistic representation that can be used to describe a figure [9]. Merin paul et al. [11] describe an approach to detect the text plagiarism. It uses semantic role labelling and sentence ranking technique to reduce the comparison delay.

3. System Architecture

Figure 1 shows the overall system structure. It consist of mainly 4 modules: Pre-processing, Shape detection, Graph creation and Comparison module.

Download English Version:

<https://daneshyari.com/en/article/6902328>

Download Persian Version:

<https://daneshyari.com/article/6902328>

[Daneshyari.com](https://daneshyari.com)