Plagarism Detection of Images

Vishwasree[1], SaraSwarnalatha[2],Arcot SaiVarshini[3],B.Mounika[4]

[1] Assistant Professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College,Hyderabad,India

[2,3,4] U.G Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College,Hyderabad,India

Corresponding author: [1]shastrivishwashree@gmail.com,[2]swarnasara99@gmail.com, [3]arcotvarshini123@gmail.com,[4]yadavmounika757@gmail.com

ABSTRACT

Tools which are designed to deal with plagiarism act on the text and generally ignore images. Plagiarism occurs when someone takes another author's work, opinions, ideas, etc. and claims them as his or her own without their consent. There has been a lot of effort done in the past to detect plagiarism in text, but there hasn't been much work done in this field.

Images transfer large amount of information in an article or scientific research. As images include wide range and a lot of flowcharts. Flowcharts carry a lot of information and this could be one of the plagarism. Image plagiarism is detected through image processing, which allows you to edit and conduct operations on images. The purpose of our project is to examine the plagiarism rate of a paper in terms of flowchart images using artificial neural network. And the main advantage of this is it is less time taking.

Keywords— *Plagarism, Artificial Neural Network*

Number: 10.14704/nq.2022.20.7.NQ33263 **Neuro Quantology 2022; 20(7):2078-2082**

Introduction

Plagiarism is defined as the act of duplicating or stealing someone else's work or ideas and claiming as one's own. Plagiarism is not just limited to photographs, architecture, flow diagrams, evenscreenshots of test results can all be copied. Students who are detected plagiarising in the classroom face a variety of sanctions and punishments, which can include expulsion. Plagiarism is not a crime in and of itself, but rather a copyright violation. Plagiarism is a serious breach of ethics in academics and other professions that are sensitive to copyright infringement. The law cannot and normally does not punish plagiarism, hence it is up to the institution to decide how to deal with it once it occurs .

Plagarism is divided into two parts, one is text based plagarism and the other one is image based plagarism. These types of plagiarism detection arecommonly used

To make a greater impact for text-matching technologies and improve detection capabilities for plagiarism, numerousapproaches have been suggested. Photographsand photo-realistic renderings are also included in our definition. Images make it possible to express a lot of information in a small amount of space, and they do so in a

different way than text does. Images are a promising factor to investigate when measuring the semantic similarity of academic writings because of these qualities. Detecting translated plagiarism and concept plagiarism requires the ability to recognise semantic similarity. If the data values can be reconstructed from graphs, even data plagiarism can be detected in some situations.

II .OBJECTIVE

Plagiarism in research is debated considerably more today than in the past. The Web's conditions, as well as the ability to conduct complicated and intelligent searches in a short period of time, are rated to this, and as a result, severe damage to research has occurred. Plagiarism detection software focuses on text and ignores visuals. Images, on the other hand, arean inseparable aspect of information transfer because they convey a great amount of data in an article or scientific research. Because graphics cover such a broad range, and because vast amounts of flowchart images were discovered in computer texts, and because flowcharts hold a lot of information, they could be one of the alternatives for plagiarism.

The goal of this paper is to use an artificial neural network to investigate a publication's plagiarism rate in terms of flowchart image plagiarism. The suggested method's average flowchart image identification accuracy in terms of structure, nodes, and edges is

81.91 percent, confirming its success.

III. LITERAURE SURVEY

In this section already existing systems are discussed.

A. Image Analysis for Plagiarism Detection:

Hurtik and Hodakova used higher degree F-transform to provide a highly efficient and reliable method to identify exact copies of photographs or cropped parts there. This method are unaffected by affine image transformations like scaling and rotation, and are relatively resistant to changes in illumination and noise. The similarity of two images can be measured by comparing their pHash value. However, the method does not consider image alterations aside from cropping.

B. Comparing Images for Document Plagiarism Detection:

This paper describes the findings of a study aimed at applying image processing algorithms to document comparisons with the goal of using them in plagiarism detection systems. Various combination of feature point detectors and descriptors are investigated as potential tool for finding similar images in the document. The methods are tested on the database consisting 5 well known image processing test images.

C. Comparing and combining Content- and Citation-based approaches for plagiarism detection:

In this paper, comparing content and citation-

based approaches with the goal of evaluating whether they are complementary and if their combination can improve the quality of the detection. They carry out experiments with real data sets of scientific papers and concluded that combination of the methods can be beneficial.

D. ImageNet Classification with Deep ConvolutionalNeural Networks:

In this paper, they trained a large, deep convolutional neural network to classify the

1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, they achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous one.

E. Reducing Computational Effort for Plagiarism Detection by using Citation Characteristics to Limit Retrieval Space

This work offers an approach to plagiarism detection in academic texts that combines citation- based, semantic argument structure, and semantic word similarity detection approaches with character-based detection methods to improve detection performance for disguise plagiarism types. Text string comparisons are the only method of plagiarism detection currently accessible. These methods uncover duplicates, but they missplagiarism that is disguised, such as paraphrases, translations, or ideas. When compared to character-based techniques, detection systems that analyse semantic similarity on a word and sentence level have consistently demonstrated superior detection accuracy for disguised plagiarism forms. The suggested hybrid strategy reduces the retrieval space with a low loss in detection accuracy by using citation-based methods as a preliminary heuristic.

IV. PROPOSED SOLUTION AND IMPLEMENTATION

This system is split into 2 phases which is the training phase and the testing phase. The training phase consists of the feature interpretation step . Inthe feature interpretation step the first thing the system does is identify whether the images are a flowchart or a non-flowchart. For the testing phase,the system will identify whether the images are a

flowchart or non-flowchart. The system will ignore

and discard images that are not flowchart.For the flowchart images, the system first pre-processes them to remove noise and then the feature extraction is done.

This system is split into 2 phases which is the training phase and the testing phase. The training phase consists of the feature interpretation step . In the feature interpretation step the first thing the system does is identify whether the images are a flowchart or a non-flowchart. For the testing phase, the system will identify whether the images are a flowchart or non- flowchart. The system will ignore and discard images that are not flowchart.For the flowchart images, the system first pre-processes them to remove noise and then the feature extraction is done

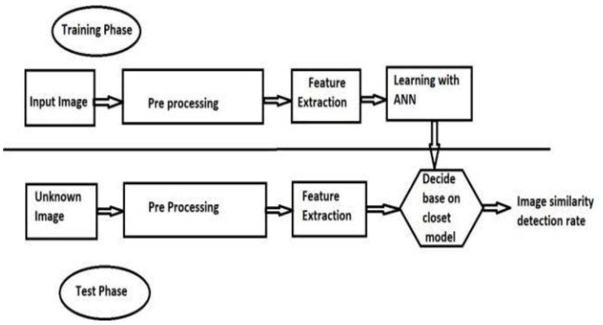

Figure 1 :Block diagram of proposed system

For input images,dimensions must be either be 364x236 or 236x364.The software accept this dimensions only and rejects the images outside of thatdimension and will give an error .User can freely input any images and here testing the software in detecting similaritites,we only picked images that already there in the database as sample images and then edited themslightly to test the accuracy of the software.

All images of database available in online and all the database of the images is used spevifically to test image retrieval software.The database contain various categories of images like buildings, plants,vehicle,animals,objects and sceneries.All of this categories we have several of 100 images ,so each of the imsge size can be considered big to test the accuracy of the software.

A . Pre-processing:-in this pre-processing stage it remove the noise in the input image so that the feature extraction can be done faster,smoother or more accurate.Sometimes features also called descriptor.The global features described the color and texture features of the entire image . Texture feature isnot always accurate since it is calculated from the

descriptor ,to extract the features of the images used by Bag-Of-Visual-Words,which is used for text-based plagiarism checker.

B. feature Extraction :- in this feature extraction ,the extracted from images and that saving into the database is color using RGB and shape using canny edge algorithm .And HSV color space,texture using tamura texture.

V. SOFTWARE DESIGN AND PROGRAMMING

SOFTWARE REQUIREMENTS:

- Operating System - Windows7/8
- Programming Language - PythonPROGRAMMING:-

UML Diagrams:

class diagram**:-**

the class diagram is the main building block of object oriented modelling.this for translating the models into programming code .In the class diagram ,classes are respresented with boxes which contain three parts :

- The upper part is for hold the name of the class
- The middle part is for attributes of the class
- The bottom part is for methods of the class

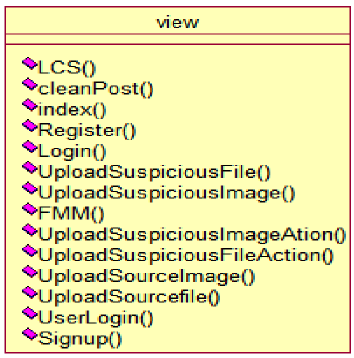

Figure 2 : class diagrammUse case diagram:-

A use case diagram can portray the different types of a system and the various ways to interact with the system.

entire image . where as in local descriptor in mainly

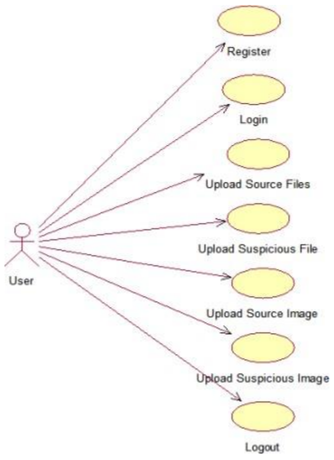

Figure 3: use case diagramActivity Diagram:

using easy-to-use, free downloadable diagramming tools. A DFD is a model for constructing and analyzing information processes. DFD illustrates the flow of information in a process depending upon the inputs and outputs. A DFD can also be referred to as a ProcessModel.

Activity diagram is a flow chart to represent the flow form one activity to another activity. The activity can be described as an operation of the system. So the control flow is drawn from one operation to another. This flow can be sequential, branched or concurrent

VI. RESULT

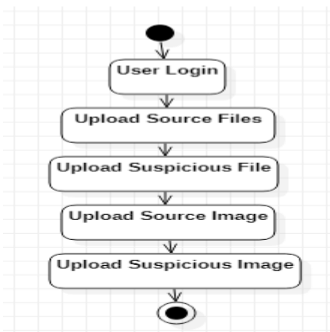

Figure 4: activity diagramData Flow Diagram:

Data flow diagrams can be used to provide a clear representation of any business function. The technique starts with an overall picture of the business and continues by analyzing each of the functional areas of interest. Data Flow Diagram (DFD) is the passage of information in a process. A DFD can be easily drawn using simple symbols. Additionally, complicated processes can be easily automated by creating DFDs

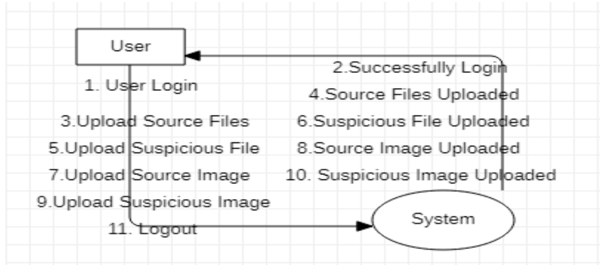

Figure 5: data flow diagram

Figure 6: Plagiarism detection

The below procedure is to be followed to detect the plagiarism:

1. Firstly install python 3.7 and then install DJANGO server and deploy code on that server and run from from browser to get above screen.

2. Then click on "New user signup".

3. Next after user signup details entered then clickon register button.

4. User signup process completed and next click on 'Login' link and enter details click on login button.

5. Click on 'Upload Source Files' link to load all files from corpus folder.

6. All files are loaded now click on 'Uploaded Suspicious File'button to load suspicious file and get result.

7. Select and uploading' angular.txt' file and then click on open button to get result and click on 'check plagiarism' button to get result.

8. Angular.txt file matched very little with corpus

file and we got similarity score as 0.03 so no plagiarism detected and now upload any file from corpus and see result.

9. Next select and uploading first file and then click on button.

10. We can see for database image and uploaded imagewe generated histogram and we can see there is no match in histogram so no plagiarismwill be detected and now close graph .

11. If histogram pixel matching score is 15173 out of 40000 pixels so image is not plagiarized.

12. At the last, the above figure shows as histrogram matching score is 40000 which means all pixels matched so plagiarism is detected in above result.

13. Similarly we can upload any text file and imageand test the application.

VII. CONCLUSION

Plagiarism detection is well known phenomenon in the academic arena. Plagiarism is copying another person's ideas, words or work and pretending they are your own something that has been copied in this way. So, Copying other people is considered as serious offence that need to be checked.The feature extracted from images and saving it to databases using the colours RGB (Red,Green,Blue) and HSV(Hue,Saturation,Value) color space. Texture using Tamura texture for the purpose of smoothnesss,roughness, contrast etc. For the shape feature extraction Canny edge algorithm is used to get the edges of the image. The result showed as 100% accuracy in case unedited images and variated accuracy in case flipped,cropped,rotated etc..

VIII. REFERENCES:

[1] Salha Alzahrani, Vasile Palade, Naomie Salim, and Ajith Abraham. 2011. Using Structural Information andCitation Evidence Detect Significant Plagiarism Cases in Scientific Publications. JASIST 63(2) (2011).

[2] Salha M. Alzahrani, Naomie Salim, and Ajith Abraham. 2012. Understanding Plagiarism Linguistic Patterns, Textual Features, and Detection Methods. In IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., Vol. 42.

[3] Yaniv Bernstein and Justin Zobel. 2004. A Scalable System for Identifying Coderivative Documents. In Proc. SPIRE. LNCS, Vol. 3246. Springer.

[4] Teddi Fishman. 2009. "We know it when we see it"?is not good enough: toward standard definition of plagiarism that transcends theft, fraud, and copyright. In Proc. Asia Pacific Conf. on Educational Integrity.

[5] Bela Gipp. 2014. Citation-based Plagiarism Detection - Detecting Disguised and Cross-language Plagiarism using Citation Pattern Analysis. Springer.

[6] Cristian Grozea and Marius Popescu. 2011. The Encoplot Similarity Measure for Automatic Detection of Plagiarism. In Proc. PAN WS at CLEF.

[7] Azhar Hadmi, William Puech, Brahim Ait Es Said, and Abdellah Ait Ouahman. 2012. Watermarking. Vol.2. InTech, Chapter Perceptual Image Hashing.

[8] Petr Hurtik and Petra Hodakova. 2015. FTIP: A tool for an image plagiarism detection. In Proc. SoCPaR.

[9] Marcin Iwanowski, Arkadiusz Cacko, and Grzegorz Sarwas. 2016. Comparing Images for Document Plagiarism Detection. In Proc. ICCVG.

[10] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In Proc. Multimedia.

[11] H.F. Judson. 2004. The Great Betrayal: Fraud in Science. Harcourt.