INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 464-472

e-ISSN : 2583-1062

Impact Factor : 5.725

TEXT AND IMAGE PLAGIARISM DETECTION - REVIEW

Baipalli Parul Raj[1], Yerupalli. Sanyasi Rao[2]

[1,2]Department of Master of Computer Science Miracle Educational Society Group of Institutions Vizianagram– 535216 (AP) India

ABSTRACT

Today, much more than in the past are discussed of plagiarism in the research. Conditions of the Web and Possibility of complex and smart searches in a short time, is rated to this, and as a result has arrived significant damages to the research. Tools designed to deal with plagiarism act on the text and ignore images. On the other, an inseparable part of information transfer are images that transfer the large volume of information in an article or scientific research. Because of the images include a very wide range and especially found large amounts of images in the computer's texts, and as respects, flowcharts are carrying a lot of information, could be one of the options of plagiarism. The purpose of this project is examining the plagiarism rate of a paper in terms of images plagiarism using Histogram Model.

1. INTRODUCTION

The issue of plagiarism is often discussed in the educational community across the world. It relates to the act of taking another person's work or ideas and passing it off as your own without giving credit to the original writer. Basically it reproduce the existing information in a modified format. Plagiarism is defined by S. Hannabuss as "is the act of imitating or copying or using somebody else's creation or idea without permission and presenting it as one's own [5]. Today with the huge popularity of internet, so many documents are freely accessible. Now internet is a extensive source to collect data. People can easily get their required information or data from internet and make their copy instead of writing their own text document. As recent trends show, the detection of plagiarism becomes more important as it is very easy for a plagiarist to find an appropriate text fragment that can be copied. On the other side it becomes increasingly difficult to correctly identify plagiarized sections due to the large amount of possible sources[7]. Plagiarism cases are an everyday topic, for example, in academics, journalism, scientific research and even in politics. This approach to plagiarism detection is especially useful when no reference collection is available or not all the possible copy sources are present, thus document to-document comparison algorithms cannot be used. Plagiarism is of various types like literal, integral, intrinsic, extrinsic, exact copy, text manipulation etc [3]. Similarly various plagiarism detection methodologies are present to detect plagiarism. Presently systems which are based on the text manipulation technique are not accurate enough for practical applications. Therefore, we have proposed a new easy method which is based on the text identification technique through file transfer method which uses a machine learning approach in order to detect plagiarism between text sets. It compares two files and identify how many words are similar between two files then we calculate a percentage value according to our threshold value required to detect plagiarism, through which we can get the plagiarized text series.

2. LITERATURE SURVEY

The main aim of this research work is to classify the emotional expression from the mouth region of the human face. As the initial task is to extract the mouth region from the facial image, a survey on various existing research works to segment the face expression images is reviewed and discussed.

3. PROPOSED SYSTEM

The proposed system has two phases: training and testing. They are seen as in train phase used of Histogram in learning stage and in the test phase in the recognition stage taken help from the modelling done by this network. Data analysis method and input image similarity detection rate with images in the database is based on the query image correlation rate with each test images and select images with the highest correlation. Correlation levels obtained at this stage report as the tested image plagiarism and the final interpretation is the responsibility of the expert.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 07, July 2023, pp : 464-472

e-ISSN : 2583-1062

Impact Factor : 5.725

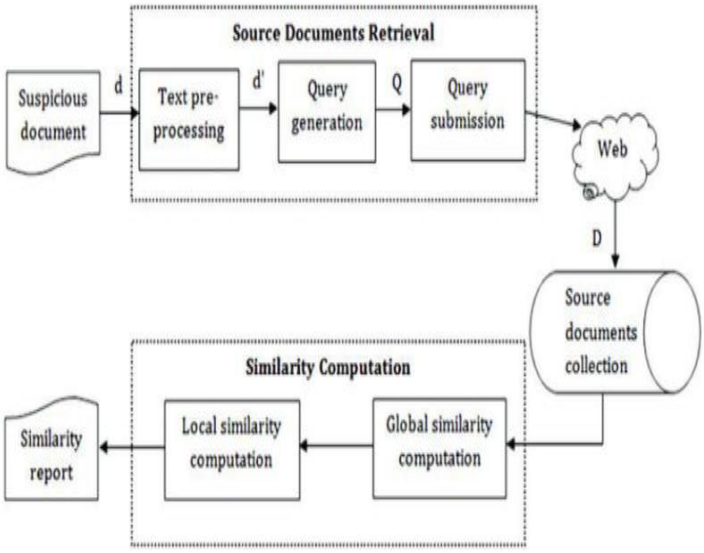
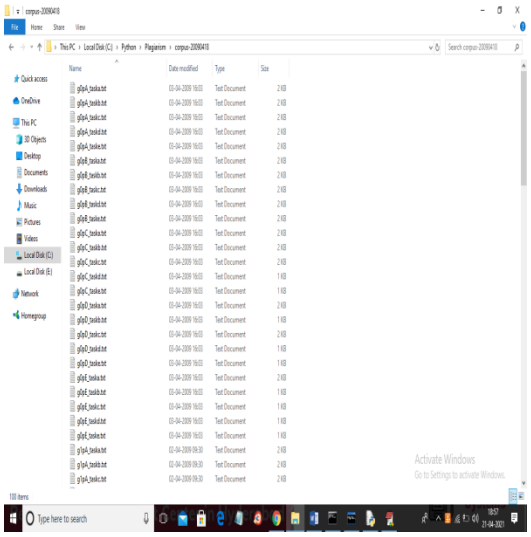
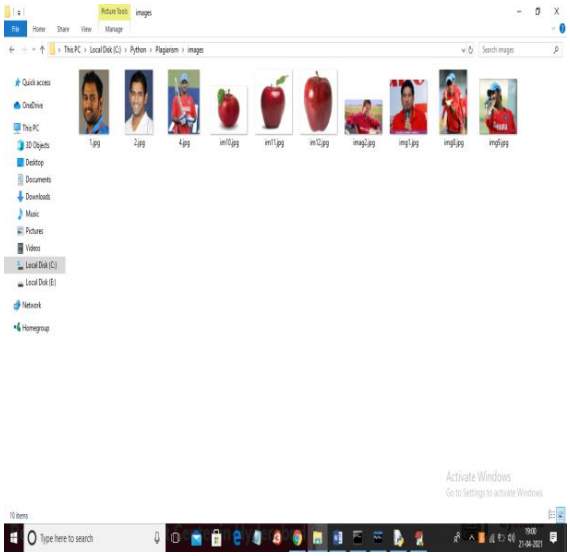


Fig. 1 Proposed System architecture

4. RESULTS



We are using below images to build histogram model and if any suspicious image similarity finds with this histogram then plagiarism will be detected. See below images used to build histogram model



Above images are available inside "images" folder

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

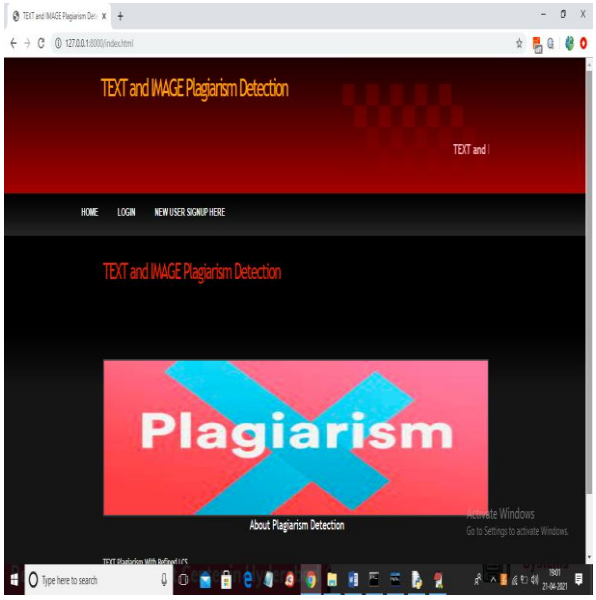
Vol. 03, Issue 07, July 2023, pp : 464-472

e-ISSN : 2583-1062

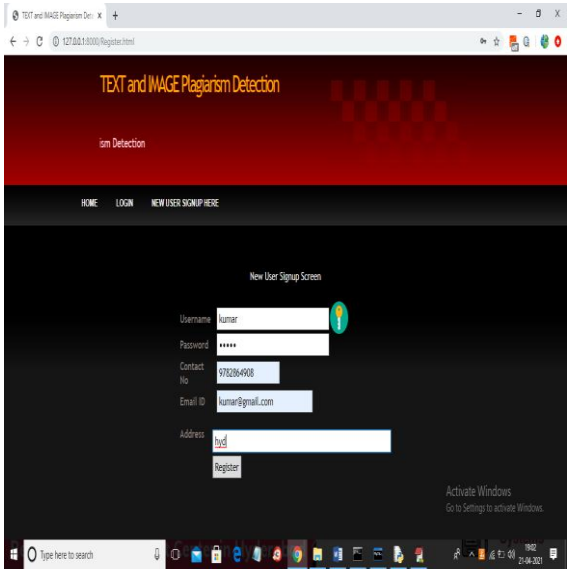
Impact Factor : 5.725

5. SCREEN SHOTS

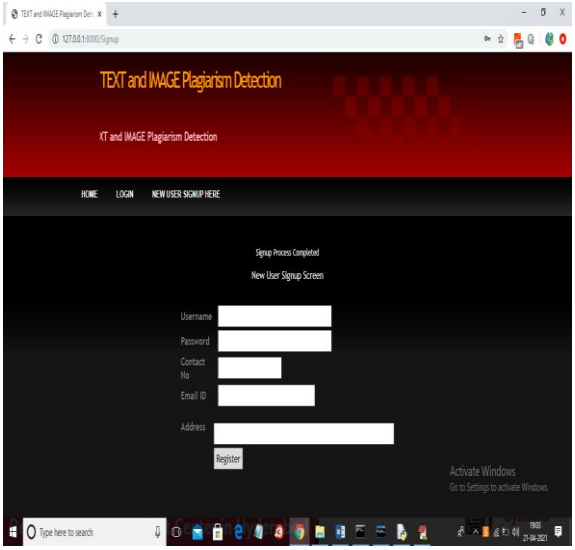
To run project install python 3.7 and then install DJANGO server and deploy code on that server and run from browser to get below screen



In above screen click on 'New User Signup Here' link to get below screen



In above screen user signup details entered and then click on 'Register' button to get below screen



www.ijprems.com
editor@ijprems.com

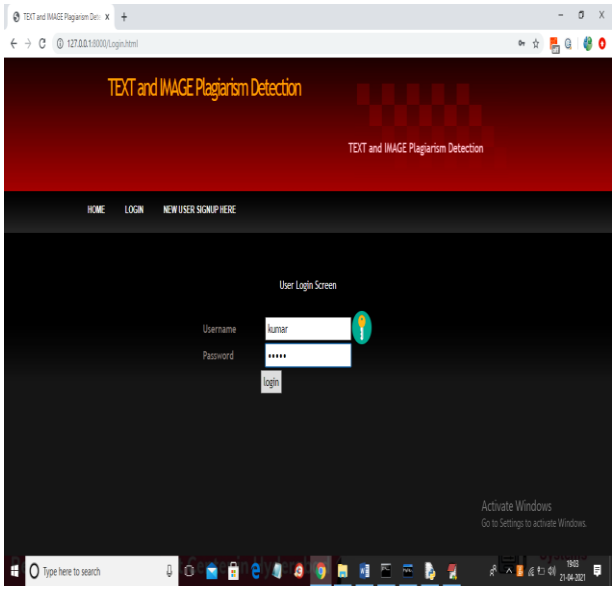
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 07, July 2023, pp : 464-472

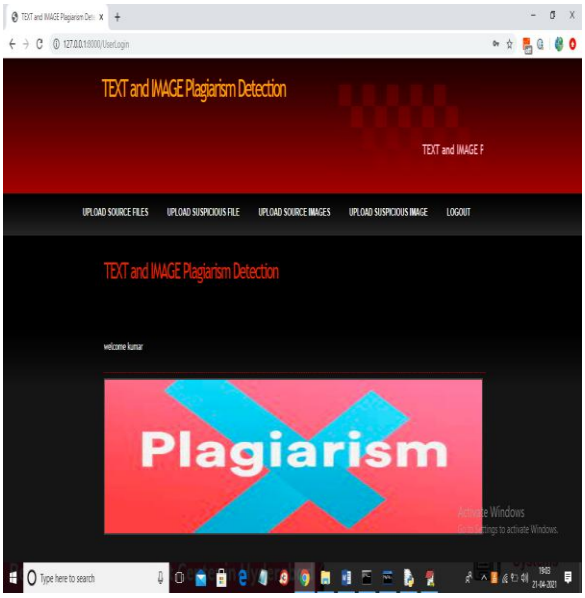
e-ISSN : 2583-1062

Impact Factor : 5.725

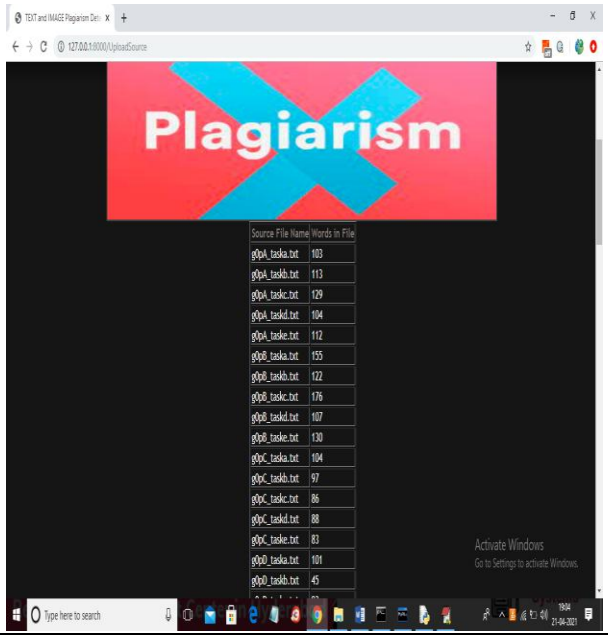
In above screen user signup process completed and now click on 'Login' link to get below screen



In above screen user is login and then click on button to get below screen



In above screen click on 'Upload Source Files' link to load all files from corpus folder



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

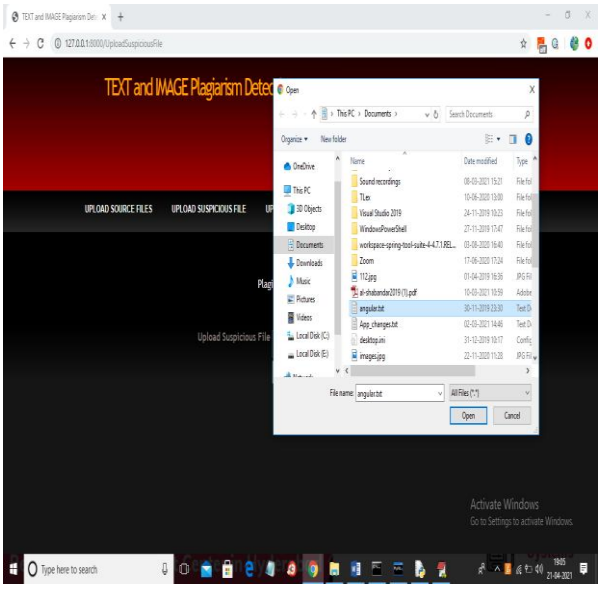
www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 464-472

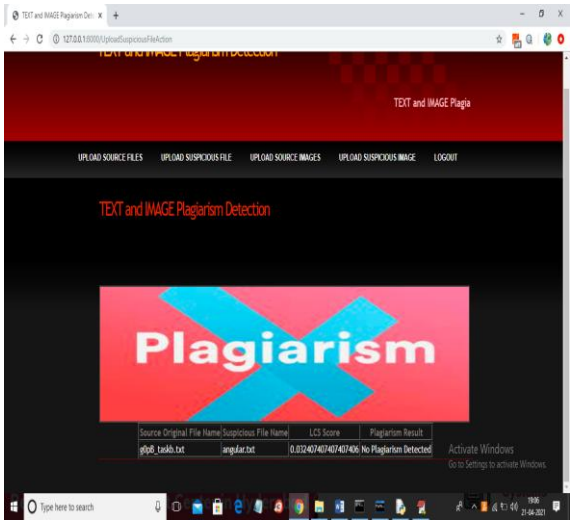
e-ISSN : 2583-1062

Impact Factor : 5.725

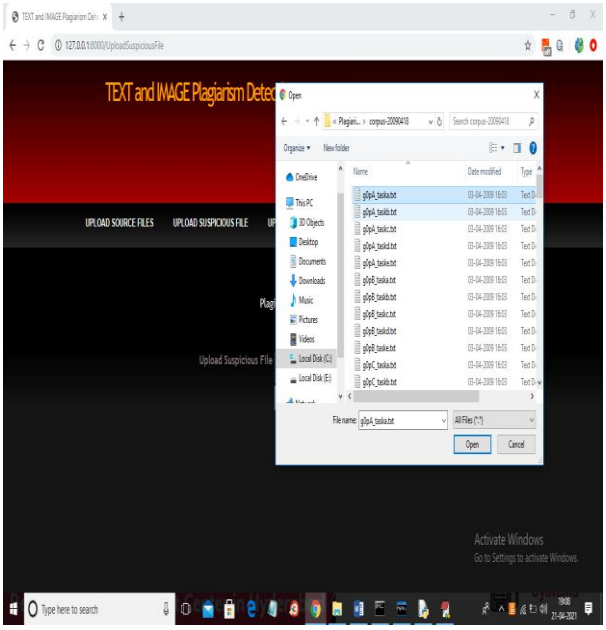
In above screen all files are loaded now click on 'Upload Suspicious File' button to load suspicious file and get result



In above screen I am selecting and uploading 'angular.txt' file and then click on 'Open' button to get below result and then click on 'Check Plagiarism' button to get result



In above screen angular.txt file matched very little with g)pB_taskb.txt corpus file and we got similarity score as 0.03 so no plagiarism detected and now upload any file from corpus and see result



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

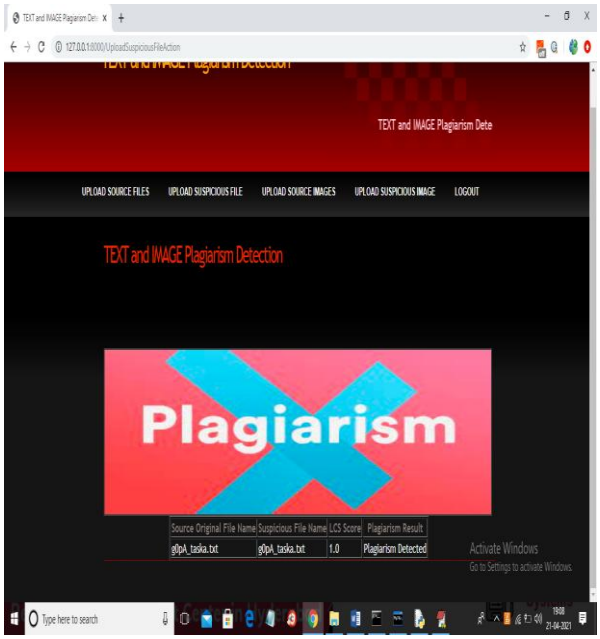
www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 464-472

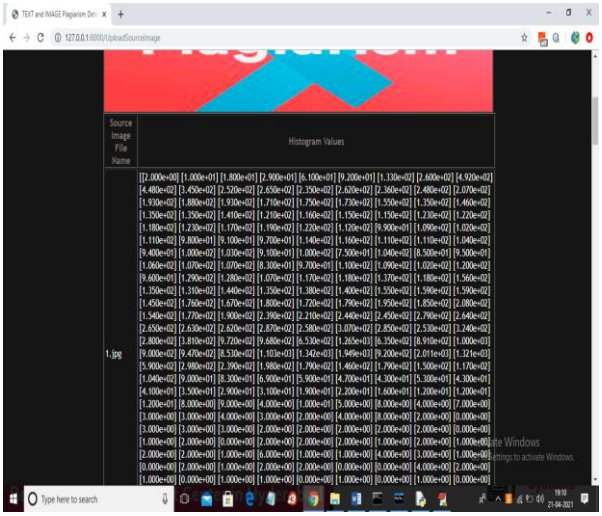
e-ISSN : 2583-1062

Impact Factor : 5.725

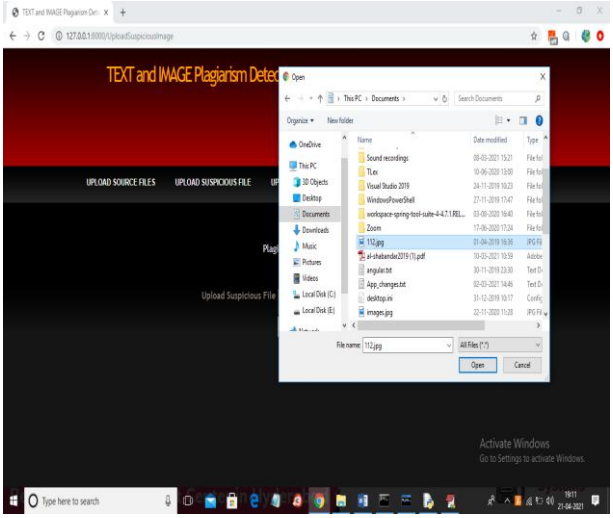
In above screen I am selecting and uploading first file and then click on button to get below result



In above screen LCS score is 1.0 which means 100% matched with corpus file so plagiarism detected and similarly not only this u may enter any text file and get result. Now click on 'Upload Source Images' link to upload all images from 'images' folder



In above screen from all database images histogram will be calculated and store in array and whenever we upload new test image then both histogram will get matched and now click on 'Upload Suspicious Image' link to upload some image



www.ijprems.com
editor@ijprems.com

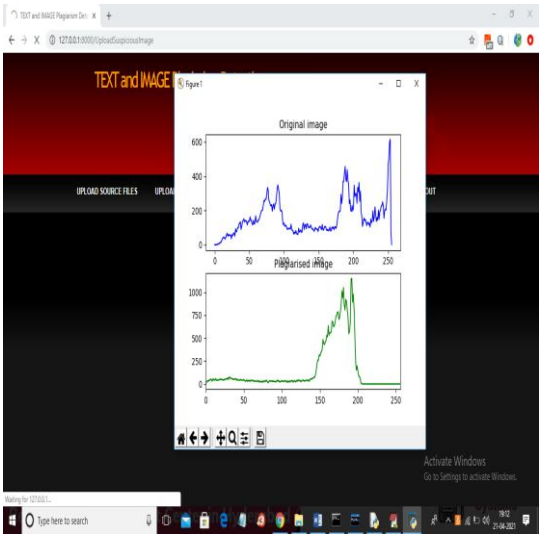
INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 07, July 2023, pp : 464-472

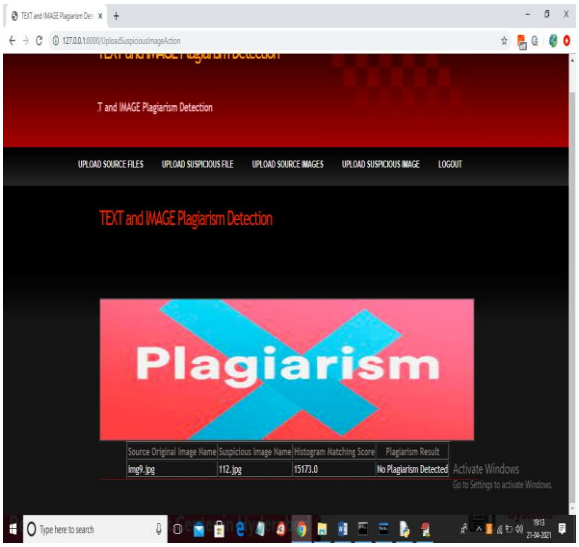
e-ISSN : 2583-1062

Impact Factor : 5.725

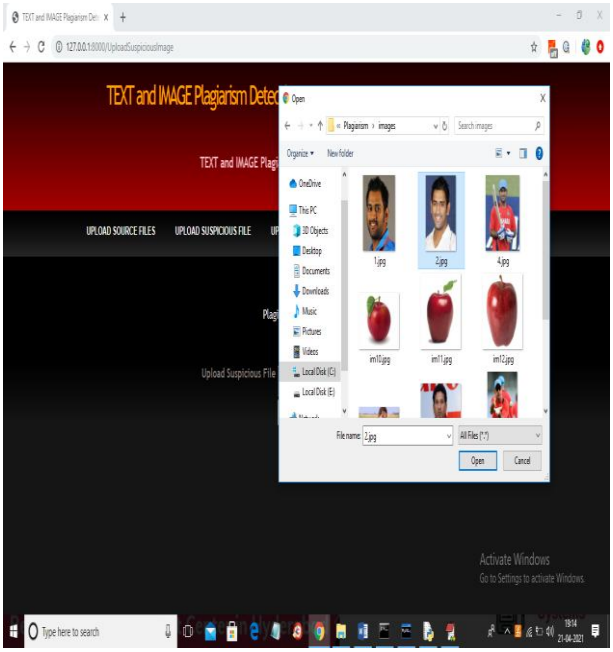
In above screen I am selecting and uploading '112.jpg' file and then click on 'Open' button to get below result



In above screen we can see for database image and uploaded image we generated histogram and we can see there is no match in histogram so no plagiarism will be detected and now close above graph to get below result



In above screen histogram pixel matching score is 15173 out of 40000 pixels so image is not plagiarised and now upload image from "images" folder and see result



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

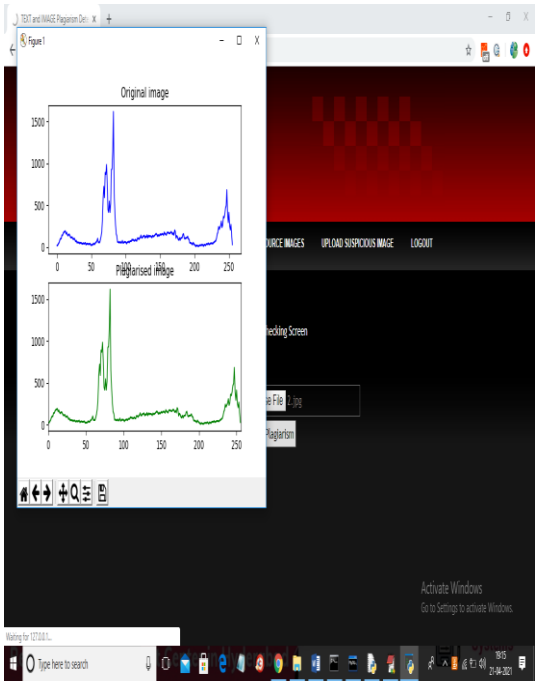
www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 07, July 2023, pp : 464-472

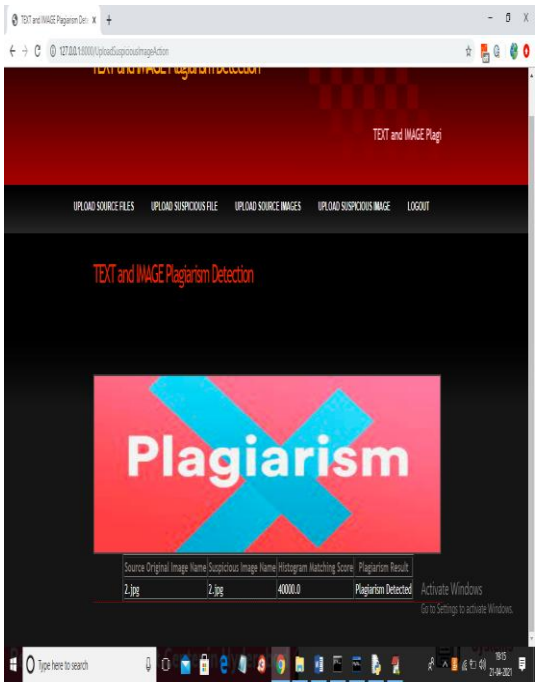
e-ISSN : 2583-1062

Impact Factor : 5.725

In above screen I am selecting and uploading '2.jpg' file from "images" database folder and below is the result



In above screen we can both original and uploaded image histogram is matching 100% so plagiarism is detected and now close above graph to get below result



In above screen histogram matching score is 40000 which means all pixels matched so plagiarism is detected in above result.

Similarly u can upload any text file and image and test the application

6. CONCLUSION

The issue of plagiarism in academic research is receiving more attention than ever. Web conditions and the capacity to do complex and sophisticated searches in a short amount of time have had a significant impact on research. Visuals are ignored by text-focused plagiarism detection programmers. When it comes to conveying the huge quantities of information included in a research paper or ot

her academic writing, images are an important part of the process. It's probable that computer-generated texts include plagiarism due to the large quantity and diversity of images available, as well as the fact that flowcharts contain a great deal of information. Our goal is to detect how many images in a paper have been plagiarized using the Histogram Model.

7. REFERENCES

[1] Imam Much Ibnu Subroto and Ali Selamat, "Plagiarism Detection through Internet using Hybrid Artificial Neural Network and Support Vectors Machine," TELKOMNIKA, Vol.12, No.1, March 2014, pp. 209-218.

[2] Upul Bandara and Gamini Wijayrathna ,"Detection of Source Code Plagiarism Using Machine Learning Approach," International Journal of Computer Theory and Engineering, Vol. 4, No. 5, October 2012, pp.674-678.

[3] SalhaAlzahrani, Naomie Salim, Ajith Abraham, and Vasile Palade," iPlag: Intelligent Plagiarism Reasoner in Scientific Publications," IEEE World Congress on Information and Communication Technologies, 2011.

[4] Barrón Cedeño, A., & Rosso, "On automatic plagiarism detection based on n-grams comparison," In Advances in Information Retrieval, Vol. 5478. Lecture Notes in Computer Science, pp. 696–700, Springer.

[5] Ahmad Gull Liaqat and Aijaz Ahmad, "Plagiarism Detection in Java Code," Degree Project, Linnaeus University, June 2011, pp. 1-7.

[6] A Selamat, IMI Subroto and Choon-Ching Ng, "Arabic Script Web Page Language Identification Using HybridKNN Method," International Journal of Computational Intelligence and Applications, 2009, pp. 315-343.

[7] Michael Tschuggnall and Gunther Specht , "Detecting Plagiarism in Text Documents through Grammar-Analysis of Authors," pp. 241-255.

[8] Bill B. Wang, R I. (Bob) McKay, Hussein A. Abbass and Michael Barlow, "Learning Text Classifier using the Domain Concept Hierarchy," ACT 2600, pp. 1-5.

[9] Francisco R., Antonio G., Santiago R., Jose L., Pedraza M., and Manuel N., ―Detection of Plagiarism in Programming Assignments, ‖ IEEE Transactions on Education, vol. 51, No.2, pp.174-183, 2008.