

Detection of Plagiarism for Flowcharts Figures with Text

Sagar A.Lingayat¹, Nayan R. Baviskar², Prof. Chetana T.Baviskar³,

*B.E. Student, Dept. of Computer Engineering, SSBT's, College of Engineering and Technology, Jalgaon, Maharashtra, India*¹

*B.E. Student, Dept. of Computer Engineering, SSBT's, College of Engineering and Technology, Jalgaon, Maharashtra, India*²

*Asst. Professor, Dept. of Computer Engineering, SSBT, College of Engineering and Technology, Jalgaon, Maharashtra, India*³

Abstract

Plagiarism is the act of taking writing, conversation, song, or even idea of someone else and passing it off as your own. While Considering the today's huge use of internet the people just copies the text from any of the paper and misuse that thing showing as their own. The current available plagiarism detection systems are text-based and code based. In text based detection system ,it simply discards the flowchart figures before detection. Discarding the flowchart figures implies that the plagiarism will remain in the text which results more possibility for plagiarism. Therefore there is a need to develop a system that will detect plagiarism in figures and flowcharts. The main goal is to create a figure plagiarism system that is primarily based on shapes which focuses on flowchart detection.

Index Terms- Plagiarism, text-based, codebased, flowchart

I. INTRODUCTION

Plagiarism is the act of representing others work as their own. Plagiarism can be considered as electronic crimes like hacking. In the last few years, Plagiarism detection is an important aspect in today's world, where huge data resources are available to do the electronic crimes. Plagiarism can be involve in different fields such as papers, art section and source code. There are several definitions of plagiarism, for example: copy another person's work then makes it as a work of your own. This process termed as plagiarism.

Flowcharts are considered as an important issue to describe the workflow of any system. They are helpful to get a better understanding of the functional aspect which is harder to understand theoretically. While designing the projects, flowcharts and figures play an important role. Method for deriving the structural design to build entire system and solving engineering problems that can be explained by using flowcharts and other types of figures in many projects. Flowchart used in design helps in dividing problem into smaller parts and managing problems easily.

Text-based plagiarism detection systems are widely used but plagiarism detection for flowcharts is not issue that talked about by researchers. Plagiarism is serious concern that has to be stop. There are many plagiarism detection systems that had been developed to perform this checks. Most of detection techniques, discard the figures and charts before checking for plagiarism. The discarding of the figures and charts results in free area that people can take benefit from. That means people can copy figures and charts easily without worrying to be detected by the current text-based plagiarism systems. Therefore there is a need to provide a supportive system that will detect plagiarism in figures, mainly flowcharts.

Section 2 describes Related Work. Analysis of Plagiarism Tools is described in Section 3. Section 4 describes Proposed System.

II. RELATED WORK

Presently, there are only a few papers that are talking about figure plagiarism, particularly diagrams and flowcharts. Although, there are very few works have been presented in flowchart plagiarism system, there are some works related to this topic, such as methods to characterize flowchart types based on the image features.

A flowchart is the way to represent information to its reader. Concurrently, the text present in the charts may be work as features of description.

Vasudevan et al. introduces a method to extract information inside flowcharts by contour processing and neural network-based optical character recognition [2]. The method analyzes flow lines and contour of the system as feature descriptors.

Though Flowchart contains text inside it to give more information, the shapes used in flowcharts has an importance to describe the process work of system. We can say that flowchart is a kind of images, thus it can be by kinds of image features. Therefore, to develop a flowchart plagiarism system, the system should be able identify these features which will give response to human query. Zhang and Lu presented techniques to represent and describe figures based on their shape features [4]. Awal et al. took a different approach to characterize flowcharts semantically using grammatical approach [3].

Mumtaz Al-Dabbagh described the Intelligent Bar Chart Plagiarism detection in documents based on the image features. Senosy Arrish and et al. introduces the Shape based plagiarism detection in flowchart [1]. Flowcharts contained shapes are self-explanatory. So, detecting shape can make understanding what writer wants to explain.

III. ANALYSIS OF PLAGIARISM TOOLS

Currently, there are very few papers which are focuses on plagiarism detection some of them are as follows with their comparisons:

A. PlagScan

PlagScan is online software used for plagiarism checking in text. PlagScan is mostly used in schools and colleges by teachers and students for checking plagiarism in different documents. PlagScan use complex algorithms for checking and analyzing uploaded document for plagiarism detection, referred up-to-date linguistic research. PlagScan database and millions of online documents are compared with the Unique signature extracted from the document's structure. So PlagScan is able to detect most of plagiarism types either directs copy and paste or words switching, which provides an accurate measurement of the level of plagiarized content in any given documents. [8]

B. Turnitin

Turnitin is the widely used plagiarism detection software which evaluates and improves student writing. It is used as the company's cloud-based service for uniqueness checking, online ranking and each review saves instructors time and provides rich feedback to students. Turnitin is one that used by many institutions and countries to manage the submission, tracking and evaluation of online student papers. Institutions license Turnitin on an annual basis. Use of Turnitin is always suggested by institutions. Assignments on Turnitin are given by the instructors in class. Turnitin currently offers interfaces and content matching for Originality Check and Grade Mark in multiple languages [5][6].

C. EVE

The Essay Verification Engine (EVE) is a desktop application. It is capable to search large number of susceptible data on Websites over Internet. EVE is based on the similarity check technique of the submitted query. EVE supports the MS Word and plain text plagiarism checking. If it finds suspect of plagiarism, the URL is saved. As soon as the search is over, the teacher is provide a report for the plagiarism and a copy of the paper showing all plagiarized data in red highlights. [5][6]

The Comparison between Plagiarism's tools are shown in following figure:

Parameter	Comparisons Between Plagiarism Tools			
	PlagScan	Turnitin	EVE	FlowCheck
Algorithm Used	Linguistic research	Similarity Technique	Similarity Technique	Similarity Technique
Document Language Support	English	Multi Language	English	English
Platform for client	Web-based	Web-based	Web-based	Web-based
Main users	Teacher/Student	Teacher/Student	Teacher/Student	Teacher/Student
File Submission Format	MS Word Doc and PDFs	MS Word Doc	MS Word Doc and Plain Text	MS Word Doc
Checks for	Text	Text	Text	Flowcharts and text
Results	Web site and Instructor report	Web site and Instructor report	Instructor report	Instructor report

Figure 3.1: Comparisons of Different Plagiarism Detection Tools with FlowCheck (Proposed System).

IV. PROPOSED WORK

The proposed work is plagiarism detection on flowcharts is based on flowchart which compares flowcharts as on their shaped and the text given in that flowchart. The text document given in that detection techniques are totally neglected only the flowchart available in that papers are considered for comparison. If match found between input flowchart and database stored flowchart then it considered as plagiarism. Otherwise it stored in database for next comparison.

PROBLEM STATEMENT

Plagiarism detection is very much important today, especially for academic & researches areas. The current available plagiarism detection systems are text-based, they simply discards the figures & flowchart before detection. Discarding the figures and charts results in look holes that people can take advantage. That means without the current plagiarism systems detecting it people can plagiarized flowcharts figures and charts easily. In current there are only few papers which talks about flowcharts plagiarism detection. Hence, for detecting plagiarism in figures and charts there is a need to develop a system. The main purpose of this project is to create a flowchart figure plagiarism system that is mainly based on Shape. This proposed system primarily focuses on plagiarism in flowcharts.

OBJECTIVES

The main primarily purpose of this paper is to detect the plagiarism for flowcharts in text and to provide

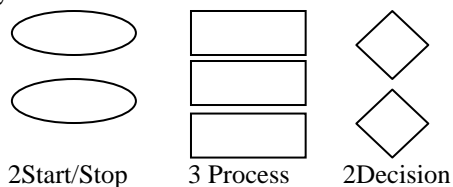
the supportive measure for the current available plagiarism detection system.

PROBLEM DEFINITION

The System should be able to check for the plagiarized flowchart. Many available plagiarism detection techniques focuses only on text but figures and flowcharts are totally neglected by these techniques. If the flowcharts are neglected to check, it leads to make vulnerability for copying flowcharts. Due to this, the data remain plagiarized. Shape base plagiarism detection is method where the plagiarism detection for flowchart can be implement more efficiently .Shape based plagiarism detection system will maintain a personalised knowledge database which will contain flowcharts, so that when user provide an input document then a comparison will be made between user input document and the flowcharts present in knowledge database and based on this comparison a report will be generated which will detect whether the input document is plagiarized or not. Here user will provide the input document, the system manager will incorporate the various functionality like comparison between the flowchart, updating of database if required and report generation and the Knowledge database will support the functionality provided by the System Manager.

STEPS FOR DETECTION OF FLOWCHART PLAGIARISM [1]

1. Load the pre-processed figure.
2. Convert to grey scale.



3. Detect edges using canny edge detection.

4. Localize the each shape in the figure.

5. For each localized shape.

a. Determine the centroid.

b. Determine the boundaries

c. Calculate the distance from centroid to boundary using Euclidean distance.

d. Let A = highest distance

Let B = lowest distance

Let C = number of pixel in each shape

e. Circle = $A - B$

f. Ellipse = $C / A * B * \pi$

g. Rectangle = $C / (4 * B (A^2 - B^2)^{0.5})$

h. Diamond = $(C * (A^2 - B^2)^{0.5}) / 2 * A^2 * B$

i. If {(circle < 10)

Shape = circle

Else if (ellipse < 1.05) and (ellipse > 0.95)

Shape = ellipse

Else if (rectangle < 1.05) and (rectangle > 0.95)

Shape = rectangle

Else if (diamond < 1.05) and (diamond > 0.95)

Shape = diamond }

$$Q = 0h + 2i + 3j + 2k$$

Where, dimension represents

h: Connector Shape

i: Start/Stop Shape

j: Process Shape

k: Decision Shape

The detected shapes from the user query are counted and inserted in above formula of getting 'q', which is a threshold value that compared with the document vector in the database to check the similarity.

The text in the query flowchart is compared with text in flowcharts that are stored in the database. The check is based on text similarity. The rank is calculated on the basis of this similarity of text and similarity in flowchart shapes. On these aspect plagiarism is decided.

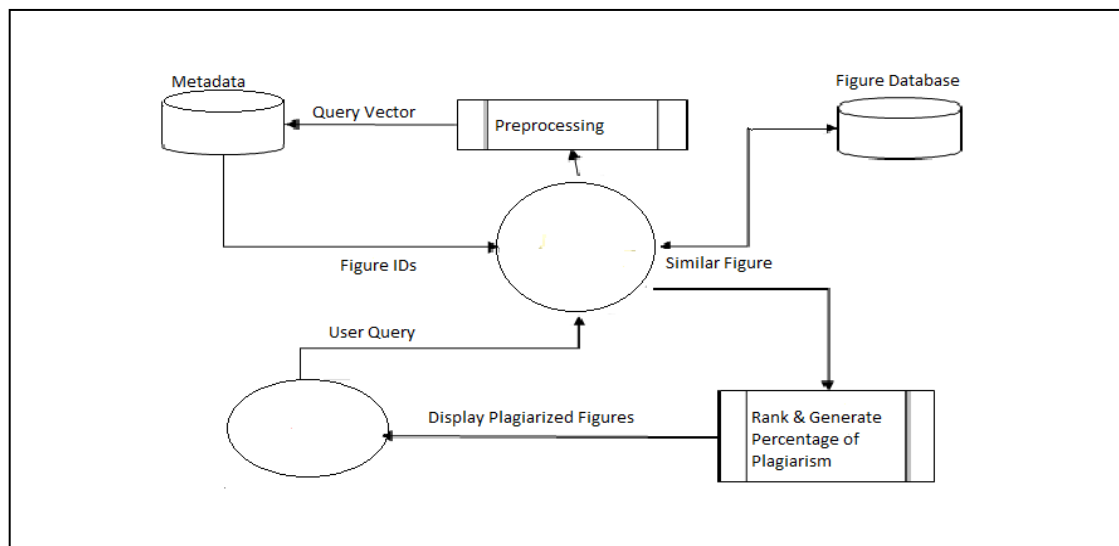


Figure 4.2: Block Diagram for Flowchart Plagiarism Detection Process

The most important part of the figure plagiarism detection system is the plagiarism retrieval search engine which described in Figure 4.2. The user query is treated as a sample figure by the search engine. The sample figure is pre-processed to derive the query vector that will be

compared with the figure-document vectors on the metadata database. Similarity between the query vector and each figure-document is calculated using the cosine similarity check. The figure id, ranked of the most alike figures are fetched from the database and reported to the user.

V. CONCLUSION

The proposed system is going to develop in such way that it is able to retrieve figures with their certain characteristics. These retrieved figures are used to find out the similarities with the prepared flowcharts of authorized person. The proposed system will utilize Canny Edge detection algorithm to find out the edges of the shapes which is a successful algorithm in image processing. The similarity checking between user query and the documents stored in database is based on the figure id and ranks which will give fine results. Hence, the proposed system can be helpful in such direction.

REFERENCES

- [1] Senosy Arrish, Fadhil Noer Afif, Ahmadu Maidorawa and Naomie Salim, "Shape-Based Plagiarism Detection for Flowchart Figures in Texts" Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Malaysia, International Journal of Computer Science & Information Technology (IJCSIT) Vol 6, No 1, February 2014 DOI :
- [2] Vasudevan, B.G., Dhanapanichkul, S., Balakrishnan, R., "Flowchart knowledge extraction on image processing," in IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). Hong Kong, 2008, pp. 4075-4082.
- [3] Ahmad-Montaser Awal, Guihuan Feng, Harold Mouchère, Christian Viard-Gaudin, "First experiments on a new online handwritten flowchart database," Document Recognition and Retrieval XVIII, vol. 7874, p. 10, 2011.
- [4] D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern Recognition, vol. 37, pp. 1-19, 2004.
- [5] Alaa m. Riad, farahat f. Farahat, aziza s. Asem & mahmoud a. Zaher, "Studying different methods for plagiarism detection", International Journal of Computer Science and Engineering (IJCSE), ISSN(P): 2278-9960, Vol. 2, Issue 5, Nov 2013, 147-154
- [6] Chao Liu, Chen Chen, Jiawei Han, "GPLAG: detection of software plagiarism by program dependence graph analysis", the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. 2006, ACM: Philadelphia, PA, USA.
- [7] C. Hung-Chi, W. Jenq-Haur, and C. Chih-Yi, "Finding Event-Relevant Content from the Web Using a Near-Duplicate Detection

Approach", the IEEE/ACM International Conference on Web Intelligence. 2007, IEEE Computer Society.

[8] <http://www.plagscan.com>. (last access November 7, 2014).