IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906258 |

A Framework for Detecting Source Code Plagiarism and Grammar Checker

Onkar Ople, Vishwanath Kalshetti, Karthik Moolya, Sumit Takawale, Prof.S.B.Akhade

Department of Computer Engineering, KJ College of Engineering and Management Research, Savitribai Phule Pune University, Pune, India

ABSTRACT: Academic dishonesty is a universal problem. The educational community across the world is facing the increasing problem of plagiarism. This widespread problem has motivated the need of an efficient, robust and fast detection procedure that is difficult to be achieved manually. Plagiarism is considered as a fraudulent act which simply means forgery of someone's fresh content by not paying any tribute. It is proliferated due to increasing access to multiple resources and simply using copy paste method. With the immense growth of internet resources new technologies and original ideas or innovative thoughts are available effortlessly due to which protection over such intellectual property is challenging, but important. Therefore, in this paper we proposed a system for detecting plagiarism and checking grammar using Application Programming Interface.

I. INTRODUCTION

1.1 BACKGROUND

Plagiarism is considered as an act of gaining someone's academic knowledge illegally without providing credit or acknowledgement. With the incremental growth of various internet resources, availability of large number data is easily possible due to which plagiarism and copyright infringement occurs. Hence, it's a need to give attention on detection of plagiarism by comparing the test documents with the registered documents. It is considered as forgery or piracy of violating laws for copyright document. It is based on two attributes (1) Stealing of another person's words, texts without any acknowledgement or (2) Adapting someone's ideas, restructuring or changing its grammar style. In this scholarly time students, professors or researchers perform some common form of plagiarism activities such as using someone's work without quoting the sources. Digitalization enables easy availability of text on web interrelated to several academic areas. Due to this problem several authors rewrite the data in their documents from original sources.
Plagiarism can be seen in two varieties which depend on the basis of work done on any document or program (a) Textual Plagiarism and (b) Source code. Scamming of text can be seen where someone uses words, ideas and simply presenting in other way such as using synonyms or improving its grammar structure. Here we will concentrate on Textual Plagiarism which is easily performed in research and education.

On the basis of their characteristics, we can further categorize into two types Literal and Intelligent plagiarism. The former is comprised of Replicate, Self-plagiarism, Accidental and Mosaic plagiarism and the later one is consist of idea, Metaphor and Structural plagiarism.

II. BACKGROUND AND LITERATURE

El Mostafa Hambi et al proposed a plagiarism detection framework based on three deep learning models: Doc2vec, Siamese Long Short-term Memory (SLSTM) and Convolutional Neural Network (CNN). Our system uses three layers: Preprocessing Layer including word embedding, Learning Layers and Detection Layer. To evaluate our system, we carried out a study on plagiarism detection tools from the academic field and make a comparison based on a set of features. Compared to other works, our approach performs a good accuracy of 98.33 % and can detect different types of plagiarism, enables to specify another dataset and supports to compare the document from an internet search [1].

El Mostafa HAMBI and Faouzia Benabbou [2] mentioned many different methods used in detection of plagiarism of ideas that are based on the Deep Learning principal. This study showed us the interest of the use of deep learning in the detection of plagiarism. As a result, the extraction of characteristics without losing the sense of the document is one of their functions. Indeed, we are capable of detecting these measurements through the result of our neuron network which provide probabilities about each kind of plagiarism already trained in a learning phase. As regard the performance of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906258 |

this approach, compared to other works, this system is capable of detecting different types of plagiarism which is important in the detection of plagiarism of ideas.

The Internet of Things (IoT) is the connection through the Internet of handling devices within physical objects, allowing them to move and communicate data. These devices may be used in academia to facilitate student- instructor interaction. The instructor uploaded the questions online and students provide solutions through IoT devices on University premises. The source code similarity in diverse types of source codes, however, is hard to detect because each programming language has a specific assembly of grammar. To address this issue, a code similarity detection approach was employed to extract the similarity between different source codes. The Latent Semantic Analysis (LSA) technique was used to retrieve semantic similarity by first transforming source codes into tokens to compute and then it finding semantic similarity in a pair of tokens. The dataset contained five different source codes: C, C#, C++, Python and Java [3].

III. PROBLEM STATEMENT

Source code plagiarism is a severe problem in academia. In academia programming assignments are used to evaluate students in programming courses. Therefore checking programming assignments for plagiarism is essential. If a course consists of a large number of students, it is impractical to check each assignment by a human inspector. Therefore it is essential to have automated tools in order to assist detection of plagiarism in programming assignments.

3.1 REQUIREMENT SPECIFICATIONS

3.1.1 User classes and Characteristics

Patient/user

- Registration
- Login to application
- Browse document
- Check Plagiarism
- Check grammar
- Correct grammar
- Result

Database

- Add new document
- Store different documents of literature

3.1.2 Assumptions and dependencies

- It is assume perfect working conditions with min 2 GB RAM and above 1GHz speed
- We assume that database is sufficient to handle minimal no. of test object

IV. ARCHITECTURE

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". Today with the huge popularity of internet, so many documents are freely accessible. Now internet is an 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. We propound a technique to identify plagiarism and correct grammar in any document.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906258 |

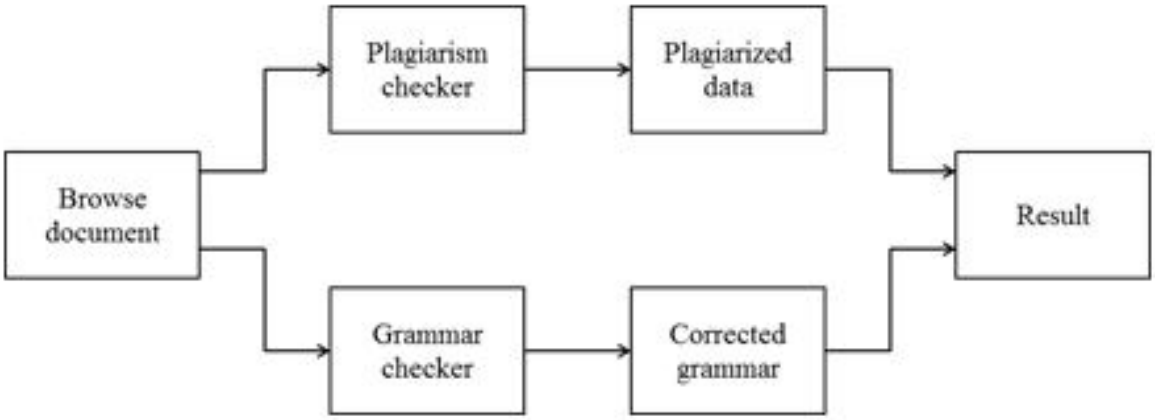


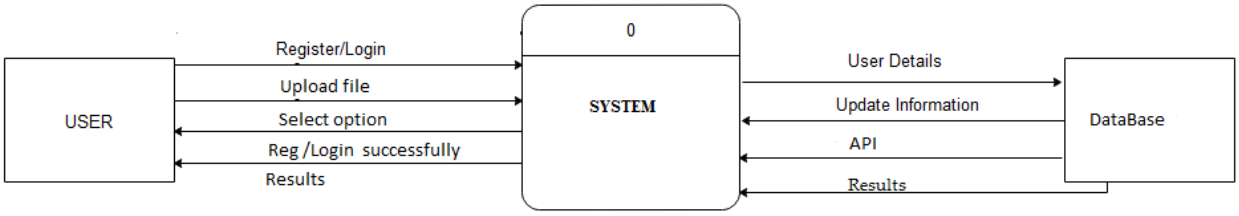
Fig block diagram of proposed system

Proposed system is developed in Python using API (Application Programming Interface). The system is divided into two parts 1) plagiarism checker and 2) grammar checker. As discussed in introduction, plagiarism is a theft which should not be neglected. Plagiarism checker scan sentences from document and compares them with different literatures present in internet. The plagiarized data is then displayed on webpage. In the end system displays plagiarism percentage and website/literature from which it is taken.

Grammar is important part of writing any idea, good grammar helps you communicate clearly and get what you want. It says "The better the grammar, the clearer the message, the more likelihood of understanding the message's intent and meaning". Non-English peoples are not that good in grammar, even if their idea is best they find it hard to convey message's intent. Hence we have developed a grammar checker for checking grammar and correcting it. In this system, grammar checker scans the document and list out all lines/sentences where grammar is wrong. Corrected grammar is displayed on webpage. Grammar corrected document is put on webpage for downloading.

4.1 DFD AND UML

DFD



V. RESULTS AND EVALUATION

5.1 TESTING STRATEGY

Testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs. Software testing can also be stated as the process of validating and verifying that a software program or application or product:

- Meets the business and technical requirements that guided its design and development;
- Works as expected; and
- Can be implemented with the same characteristics.

Software testing, depending on the testing method employed, can be implemented at any time in the development process. However, most of the test effort occurs after the requirements

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906258 |

VI. CONCLUSION

Plagiarism is a term which comes with the similar content publishing or making over the internet. Many researches outperformed in the past which need always a unique content of publishing. But the original research is also used by different researcher to use in personal content. On considering the influence of plagiarism occurring due to availability of various web articles or it is represented in number of ways. In this paper, we proposed a new system for the detection of plagiarism and grammar checking based on the deep learning methods.
In future, we will have consolidated our approach with different languages; this allows us to develop our approach to be more efficient.

REFERENCES

[1] El Mostafa Hambi, Faouzia Benabbou, "A New Online Plagiarism Detection System based on Deep Learning", (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 11, No. 9, 2020 470 | P a g e www.ijacsa.thesai.org

[2] El Mostafa Hambi, Faouzia Benabbou, "A Multi-Level Plagiarism Detection System Based on Deep Learning Algorithms", IJCSNS International Journal of Computer Science and Network Security, VOL.19 No.10, October 2019 Manuscript received October 5, 2019 Manuscript revised October 20, 2019

[3] Hamza Aldabbas, "An IOT based framework for students' interaction and plagiarism detection in programming assignments", Journal of Theoretical and Applied Information Technology 30th September 2019. Vol.97. No 18 © 2005 – ongoing JATIT & LLSISSN: 1992-8645,

[4] Anton Belyy1 Marina Dubova2 Dmitry Nekrasov "Improved Evaluation Framework for Complex Plagiarism Detection"

[5] Mausumi Sahu, "Plagiarism Detection Using Artificial Intelligence Technique In Multiple Files", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 5, ISSUE 04, APRIL 2016 ISSN 2277-8616 111 IJSTR©2016 www.ijstr.org

[6] Imtiaz Hussain Khan, Muazzam Ahmed Siddiqui, Kamal Mansoor Jambi and Abobakr Ahmed Bagais "A FRAMEWORK FOR PLAGIARISM DETECTION IN ARABIC DOCUMENTS", pp. 01–09, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.50201

[7] Ali, W., Ahmad, T., Rehman, Z., Rehman, A. U., Shah, M. A., Abbas, A., & Dustgeer, G. "A Novel Framework for Plagiarism Detection: A Case Study for Urdu Language", 2018 24th International Conference on Automation and Computing (ICAC). doi:10.23919/iconac.2018.8749122

[8] Raj Kumar Kundu and Karthik. K – "Contextual plagiarism detection using latent semantic analysis", International Research Journal of Advanced Engineering and Science, Volume 2, Issue 1, pp. 214-217, 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details