Department of Computer Science and Engineering
St. Thomas College of Engineering and Technology
Mattannur

Online Assignment Plagiarism Checker Project using Data Mining

GROUP 8

Author :
ARUNDATH (STM19CS018)
FASNA ABDUL LATHEEF (STM19CS020)
MARIYAM LIYANA (STM19CS027)
MIDHUN MOHAN B (STM19CS028)

Supervisor:
SHYAMILI PV

Academic Year
2021-2022
May 25,2022

OUTLINE

1. Outline
2. Introduction to Domain
3. Problem Identification
4. Proposal
5. Conclusion
6. References

INTRODUCTION TO DOMAIN

- **In today's digital era, data has become the most important tool.**
- **Data Mining is the process of extracting the most useful information from lots of data to quickly identify all the present trends and patterns for businesses and huge firms to understand customers and make out important decisions is called Data Mining.**
- **In simple terminology, data mining is a way to recognize hidden patterns from the extracted information of the data required for the business with the help of data wrangling techniques to categorize important data stored in proper data warehouses with the help of data mining algorithms to generate maximum revenue for a business.**

PROBLEM IDENTIFICATION

- Plagiarism is a form of academic misconduct in which you represent someone else's words or ideas as your own. The basic expectation in every class is that whatever you write will be your own words, generated from your own understanding.
- Plagiarism affects the education quality of the students and thereby reduce the economic status of the country.
- Students work less and develop a mentality to use others ideas other than their own . Use of their own talents are very low. Moreover teachers find it difficult to go through every single document to find out copied ones.

PROPOSAL

- Plagiarism is done by paraphrased works and the similarities between keywords and verbatim overlaps, change of sentences from one form to other form, which could be identified using WordNet etc.
- This plagiarism detector measures the similar text that matches and detects plagiarism. Internet has changed the student's life and also has changed their learning style. It allows the students to get deeper in the approach towards learning and making their task easier.
- Many methods are employed in detecting plagiarism. Usually plagiarism detection is done using text mining method. In this plagiarism checker software, user can register with their basic registration details and create a valid login id and password.
- By using login id and password, students can login into their personal accounts. After that students can upload assignment file, which will further divide into content and reference link. This web application will process the content, visit each reference link, and scan the content of that webpage to match the original content. Also, students can view the history of their previous documents. Students can also check the grammar mistakes on the content.

CONCLUSION

- Plagiarism detection is essential for protecting the written work. It is concluded that all institutues and and teachers should be aware of plagiarism and anti-plagiarism softwares.
- We have designed a simple method which assists us with the detection of instances of plagiarism in assignment of school and college students. Our scheme is easy to adapt for the large variety of programming languages in use, and is sufficiently robust to be highly effective in an educational environment.
- While having a detection rate as good as other more complex software, it presents its report as a simple graph, enabling large numbers of assignments to be checked quickly and efficiently. By using data mining algorithm and NLP it will provides straightforward documentation which can be used as clear and convincing evidence should a suspected instance of plagiarism be disputed.

REFERENCES

1. O. Karnalim and Simon , "Syntax trees and information retrieval to improve code similarity detection," in Proc. 32nd Australas. Comput. Edu.Conf., Feb. 2020, pp. 48 55.
2. P. Vamplew and J. Dermoudy, ``An anti-plagiarism editor for software development courses," in Proc. 7th Australas. Conf. Comput. Educ., 2005, pp. 8390.
3. “Software metrics and plagiarism detection,” J. Syst. Software, vol. 13, pp. 131–138, 1990.
4. V. Ljubovic. (2020). *Programming Homework Dataset for Plagiarism Detection*. [Online]. Available: http://dx.doi.org/10.21227/71fw-ss32

REFERENCES

5. E. Pajic and V. Ljubovic, ``Improving plagiarism detection using geneticalgorithm,"in *Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2019, pp. 571576.
6. C. Igel and M. Hüsken, ``Improving the Rprop learning algorithm,'' in *Proc. 2nd Int. Symp. Neural Comput.*, 2000, pp. 11 5121.
7. https://iopscience.iop.org/article/10.1088/1742-6596/1979/1/012070/meta
8. https://ieeexplore.ieee.org/abstract/document/9097285