



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Cheattell

Plagiarism Checker for Blogpost Articles

A dissertation by

Mr. Akila Nanayakkara

Supervised by

Mrs. Theja Perera

Submitted in partial fulfilment of the requirements for the

BSc (Hons) in Computer Science

Degree at the University of Westminster.

July 2022

Abstract

Plagiarism detection is a method of finding similar content that has been copied from another source. It is a huge problem when it comes to the search engine optimization. The problem arises when writing a blog, it needs to be 100% unique otherwise it will be detected as plagiarized content from the search engines and the SEO results of the website will go down. It is a disadvantage for the website owner or to the company and also for the user who is coming to get the information. Most of the website developing companies hire or outsource content writers to write articles for their websites. But they won't check if the articles that they are writing is unique or not. When the developers upload these articles the website because of the plagiarism levels the ranking scores will get glow. The same scenario goes with individual blog writers as well. They write blogs taking content from other websites. This research is focused on this matter, to detect blog article plagiarism before an article gets published to the internet.

To overcome this problem, Plagiarism Detecting techniques can be used to overcome this problem. Detecting plagiarism and understanding the uniqueness of the blog article before they are being published on the internet can greatly improve the blog article's SEO ranking. Also, there are many plagiarisms detector already existing, but they all are focused on academic plagiarism. The proposed system is mainly focused on blog articles and finding similar blog articles from the internet.

This research project introduces a new natural language processing based technique in order to find plagiarized content on blog articles by surfing the internet. The proposed system processes the similar findings from the internet and generates a uniqueness percentage. Depends on that we can know what are the best articles that can be published on the internet that are going to rank high in Google's search engine optimization rankings. This will make blog article writers and web developers lives much more easier.

Keywords: Plagiarism Checker, Blog Article Plagiarism Checker, Natural Language Processing